91690

STIC-Biotech/ChemLib

From: Slobodyansky, Elizabeth
Sent: Wednesday, April 16, 2003 2:30 PM
To: STIC-Biotech/ChemLib
Subject: 10/067,291

Please search for case 10/067,291:

28

SEQ ID NOs: 1 and 2 against commercial ~~and interference~~ databases.

Thank you.

CRF

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652
CM1 10D11
703-306-3222

mail box 10D01

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

Searcher:_____
Phone:_____
Location:_____
Date Picked Up: 4/17
Date Completed: 4/25
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences: 1
AA Sequences: 1
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.:_____
WWW/Internet:_____
Other (specify):_____

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:29:00 ; Search time 40 Seconds
(without alignments)
1129.299 Million cell updates/sec

Title: US-10-067-291-1
Perfect score: 1764
Sequence: 1 MTALSDLGTSNLVAVEPGAI.....RDRFWALMDEPGPLIETIEY 339

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1764	100.0	339	21	AAB26100	Mycobacterium auru
2	1759	99.7	339	21	AAB26104	Mycobacterium auru
3	847	48.0	325	19	AAW82241	A protein with ste
4	847	48.0	329	19	AAW87563	A protein with ste
5	847	48.0	330	19	AAW87564	A protein with ste
6	322.5	18.3	310	21	AAY28430	Corn branched chai
7	318.5	18.1	490	23	ABB92207	Herbicidally activ
8	315	17.9	303	18	AAW24255	Pyrobacullum aerop
9	290	16.4	282	20	AAW73375	S. aureus D-alanin
10	289	16.4	289	20	AAY13559	L. monocytogenes D
11	289	16.4	289	23	ABB49432	Listeria monocytog
12	283	16.0	282	20	AAY13560	S. haemolyticus D-
13	278.5	15.8	283	9	AAP81166	D-amino acid trans
14	278.5	15.8	283	20	AAY13562	Bacillus D-amino a
15	256	14.5	288	21	ABB06297	Heat resistant D-a
16	252	14.3	288	21	AAO20090	Protein relating t
17	250	14.2	283	17	AAW05302	D-aminotransferase
18	250	14.2	283	20	AAY13561	B. sphaericus D-am
19	250	14.2	283	23	ABB08243	Wild-type Bacillus
20	248.5	14.1	552	23	ABB93611	Herbicidally activ
21	246	13.9	283	23	ABB08244	Mutant Bacillus sp
22	245	13.9	283	23	ABP40255	Staphylococcus epi
23	244	13.8	283	19	AAW49769	Amino acid sequenc
24	238	13.5	285	21	AAY28426	Soybean branched c
25	226.5	12.8	255	21	AAY28431	Wheat branched cha
26	223.5	12.7	296	23	ABB93857	Herbicidally activ
27	222	12.6	297	21	AAY28427	Soybean branched c
28	193	10.9	174	22	ABG10774	Novel human diagno
29	165.5	9.4	358	22	AAG81462	S. epidermidis ope
30	165.5	9.4	362	23	ABP40558	Staphylococcus epi
31	163.5	9.3	348	21	AAY28423	Wheat branched cha
32	161.5	9.2	243	21	AAY28425	Rice branched chai
33	159	9.0	317	21	AAG06789	Arabidopsis thalia
34	159	9.0	317	21	AAG37156	Arabidopsis thalia
35	159	9.0	317	21	AAG37326	Arabidopsis thalia
36	159	9.0	317	21	AAG38485	Arabidopsis thalia
37	159	9.0	320	21	AAG06788	Arabidopsis thalia
38	159	9.0	320	21	AAG37155	Arabidopsis thalia
39	159	9.0	320	21	AAG37325	Arabidopsis thalia
40	159	9.0	320	21	AAG38484	Arabidopsis thalia
41	159	9.0	413	21	AAG06787	Arabidopsis thalia
42	159	9.0	413	21	AAG38483	Arabidopsis thalia
43	159	9.0	414	21	AAG37324	Arabidopsis thalia
44	154.5	8.8	307	21	AAY28421	Rice branched chai
45	154.5	8.8	366	22	AAU53181	Propionibacterium

ALIGNMENTS

RESULT 1
AAB26100
ID AAB26100 standard; Protein; 339 AA.
XX
AC AAB26100;
XX
DT 15-JAN-2001 (first entry)
XX
DE Mycobacterium aurum stereoselective transamination catalyst.
XX
KW Transamination; optically active amino compound; isomer.
XX
OS Mycobacterium aurum.
XX
PN EP1038953-A1.
XX
PD 27-SEP-2000.
XX
PF 14-MAR-2000; 2000EP-0400701.
XX
PR 19-MAR-1999; 99JP-0075511.
PR 30-MAR-1999; 99JP-0088634.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Takashima Y, Mitsuda S, Wieser M;
XX
DR WPI; 2000-629576/61.
DR N-PSDB; AAA94032.
XX
PT Novel Mycobacterium protein capable of catalyzing transamination stereo
PT selectively, useful for producing optically active amino compounds,
PT comprises a sequence of 339 amino acids

```
XX
PS   Claim 2; Page 30-32; 54pp; English.
XX
CC   The present sequence is the amino acid sequence of a protein from
CC   Mycobacterium aurum which acts as a biocatalyst. It is able to catalyse
CC   transamination stereoselectively to produce an optically active amino
CC   compound.
XX
SQ   Sequence   339 AA;

  Query Match             100.0%;  Score 1764;  DB 21;  Length 339;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-168;
  Matches  339;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MTALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MTALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60

Qy     61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120

Qy    121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180

Qy    181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240

Qy    241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300

Qy    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339
          |||||||||||||||||||||||||||||||||||||||
Db    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339


RESULT 2
AAB26104
ID   AAB26104 standard; Protein; 339 AA.
XX
AC   AAB26104;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Mycobacterium aurum stereoselective transamination catalyst mutant.
XX
KW   Transamination; optically active amino compound; isomer; mutant; mutein.
XX
OS   Mycobacterium aurum.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "wild-type Thr substituted by Ala"
XX
PN   EP1038953-A1.
XX
PD   27-SEP-2000.
XX
PF   14-MAR-2000; 2000EP-0400701.
XX
PR   19-MAR-1999;   99JP-0075511.
PR   30-MAR-1999;   99JP-0088634.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Takashima Y,  Mitsuda S,  Wieser M;
XX
DR   WPI; 2000-629576/61.
DR   N-PSDB; AAA94040.
XX
PT   Novel Mycobacterium protein capable of catalyzing transamination stereo
PT   selectively, useful for producing optically active amino compounds,
PT   comprises a sequence of 339 amino acids
XX
PS   Claim 3; Page -; 54pp; English.
XX
CC   The present sequence is a mutant version of the amino acid sequence of a
CC   protein from Mycobacterium aurum which acts as a biocatalyst. It is able
CC   to catalyse transamination stereoselectively to produce an optically
CC   active amino compound.
CC   Note: This sequence is not shown in the specification but is derived
CC   from the Mycobacterium aurum shown in SEQ ID NO: 1 (see AAB26100).
XX
SQ   Sequence   339 AA;

  Query Match             99.7%;  Score 1759;  DB 21;  Length 339;
  Best Local Similarity   99.7%;  Pred. No. 5.5e-168;
  Matches  338;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 MTALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60

Qy     61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120

Qy    121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180

Qy    181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240

Qy    241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300

Qy    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339
          |||||||||||||||||||||||||||||||||||||||
Db    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339


RESULT 3
AAW82241
ID   AAW82241 standard; Protein; 325 AA.
XX
AC   AAW82241;
XX
DT   26-FEB-1999  (first entry)
XX
DE   A protein with stereoselective transaminase activity.
XX
KW   Stereoselective; transaminase activity; optically active amine;
KW   R-chirality; ketone; synthetic intermediate; pharmaceutical;
KW   agrochemical.
XX
OS   Arthrobacter sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "optional"
XX
PN   WO9848030-A1.
XX
PD   29-OCT-1998.
XX
PF   20-APR-1998;   98WO-JP01814.
XX
PR   23-APR-1997;   97JP-0121732.
XX
```

```
PA   (KANF ) KANEKA CORP.
XX
PI   Hasegawa J,  Ikenaka Y,  Iwasaki A,  Kizaki N,  Matsumoto K;
PI   Ogura M,  Yamada Y;
XX
DR   WPI; 1998-583664/49.
DR   N-PSDB; AAV73146.
XX
PT   Production of optically active amine(s) from ketone(s) and an amino
PT   donor - using a transaminase derived from an Arthrobacter species.
XX
PS   Claim 22; Page 56-57; 84pp; Japanese.
XX
CC   The present sequence represents a protein with stereoselective
CC   transaminase activity. The transaminase is used to produce
CC   the optically active amines (having R-chirality) of the
CC   invention. A ketone is treated with a transaminase in the
CC   presence of a primary amine as amino donor to produce the
CC   amines. The chiral amines are synthetic intermediates for
CC   use as pharmaceuticals and agrochemicals.
XX
SQ   Sequence   325 AA;

  Query Match            48.0%;  Score 847;  DB 19;  Length 325;
  Best Local Similarity   51.0%;  Pred. No. 2.2e-76;
  Matches  158;  Conservative   56;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      30 IQYSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRL 89
           | ||||||| ::| ||| ||||| ::|  ||:||||| |: |||:|||| |||:|| |||
Db      16 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRL 75

Qy      90 GDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDL 149
             ||::||   |  :|:    ::||: ||  : |: ::|||:||::::|||||    ||:|:
Db      76 DDHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITRGYSSTPGERDI 135

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASF 209
           :|   |||:||:|| |  |       |  |:| : |||  |:::|| :||:|||||  |
Db     136 TKHRPQVYMYAVPYQWIVPFDRIRDGVHAMVAQSVRRTPRSSIDPQVKNFQWGDLIRAVQ 195

Qy     210 EAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGI 269
           |  |||    :||| |  :||| |||| ::||| : || | ||||||||| |:|: :|
Db     196 ETHDRGFEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTVLEIAESLGH 255

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
           || | |:|  || ||||::  |||||| |  |:||  : :| |||:| :|  |:| |  |
Db     256 EAILADITLAELLDADEVLGCTTAGGVWPFVSVDGNPISDGVPGPITQSIIRRYWELNVE 315

Qy     330 PGPLIETIEY 339
             |:  ::|
Db     316 SSSLLTPVQY 325


RESULT 4
AAW87563
ID   AAW87563 standard; Protein; 329 AA.
XX
AC   AAW87563;
XX
DT   26-FEB-1999  (first entry)
XX
DE   A protein with stereoselective transaminase activity.
XX
KW   Stereoselective; transaminase activity; optically active amine;
KW   R-chirality; ketone; synthetic intermediate; pharmaceutical;
KW   agrochemical.
XX
OS   Arthrobacter sp.
XX
PN   WO9848030-A1.
XX
PD   29-OCT-1998.
XX
PF   20-APR-1998;   98WO-JP01814.
XX
PR   23-APR-1997;   97JP-0121732.
XX
PA   (KANF ) KANEKA CORP.
XX
PI   Hasegawa J,  Ikenaka Y,  Iwasaki A,  Kizaki N,  Matsumoto K;
PI   Ogura M,  Yamada Y;
XX
DR   WPI; 1998-583664/49.
DR   N-PSDB; AAV83675.
XX
PT   Production of optically active amine(s) from ketone(s) and an amino
PT   donor - using a transaminase derived from an Arthrobacter species.
XX
PS   Claim 22; Page 56-57; 84pp; Japanese.
XX
CC   The present sequence represents a protein with stereoselective
CC   transaminase activity. The transaminase is used to produce
CC   the optically active amines (having R-chirality) of the
CC   invention. A ketone is treated with a transaminase in the
CC   presence of a primary amine as amino donor to produce the
CC   amines. The chiral amines are synthetic intermediates for
CC   use as pharmaceuticals and agrochemicals.
XX
SQ   Sequence   329 AA;

  Query Match            48.0%;  Score 847;  DB 19;  Length 329;
  Best Local Similarity   51.0%;  Pred. No. 2.2e-76;
  Matches  158;  Conservative   56;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      30 IQYSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRL 89
           | ||||||| ::| ||| ||||| ::|  ||:||||| |: |||:|||| |||:|| |||
Db      20 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRL 79

Qy      90 GDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDL 149
             ||::||   |  :|:    ::||: ||  : |: ::|||:||::::|||||    ||:|:
Db      80 DDHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITRGYSSTPGERDI 139

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASF 209
           :|   |||:||:|| |  |       |  |:| : |||  |:::|| :||:|||||  |
Db     140 TKHRPQVYMYAVPYQWIVPFDRIRDGVHAMVAQSVRRTPRSSIDPQVKNFQWGDLIRAVQ 199

Qy     210 EAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGI 269
           |  |||    :||| |  :||| |||| ::||| : || | ||||||||| |:|: :|
Db     200 ETHDRGFEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTVLEIAESLGH 259

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
           || | |:|  || ||||::  |||||| |  |:||  : :| |||:| :|  |:| |  |
Db     260 EAILADITLAELLDADEVLGCTTAGGVWPFVSVDGNPISDGVPGPITQSIIRRYWELNVE 319

Qy     330 PGPLIETIEY 339
             |:  ::|
Db     320 SSSLLTPVQY 329


RESULT 5
AAW87564
ID   AAW87564 standard; Protein; 330 AA.
XX
AC   AAW87564;
XX
DT   26-FEB-1999  (first entry)
XX
DE   A protein with stereoselective transaminase activity.
XX
KW   Stereoselective; transaminase activity; optically active amine;
KW   R-chirality; ketone; synthetic intermediate; pharmaceutical;
KW   agrochemical.
XX
OS   Arthrobacter sp.
XX
PN   WO9848030-A1.
XX
```

PD 29-OCT-1998.
XX
PF 20-APR-1998; 98WO-JP01814.
XX
PR 23-APR-1997; 97JP-0121732.
XX
PA (KANF) KANEKA CORP.
XX
PI Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;
PI Ogura M, Yamada Y;
XX
DR WPI; 1998-583664/49.
DR N-PSDB; AAV83676.
XX
PT Production of optically active amine(s) from ketone(s) and an amino
PT donor - using a transaminase derived from an Arthrobacter species.
XX
PS Claim 22; Page 56-57; 84pp; Japanese.
XX
CC The present sequence represents a protein with stereoselective
CC transaminase activity. The transaminase is used to produce
CC the optically active amines (having R-chirality) of the
CC invention. A ketone is treated with a transaminase in the
CC presence of a primary amine as amino donor to produce the
CC amines. The chiral amines are synthetic intermediates for
CC use as pharmaceuticals and agrochemicals.
XX
SQ Sequence 330 AA;

Query Match 48.0%; Score 847; DB 19; Length 330;
Best Local Similarity 51.0%; Pred. No. 2.2e-76;
Matches 158; Conservative 56; Mismatches 96; Indels 0; Gaps 0;

```
Qy      30 IQYSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRL 89
           | ||||||| ::| ||| ||||| ::|  ||:||||| |: |||:|||| |||:|| |||
Db      21 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRL 80

Qy      90 GDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDL 149
            ||::||   |  :|:    ::||: ||  : |: ::|||:||::::|||    ||:|:
Db      81 DDHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITRGYSSTPGERDI 140

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASF 209
           :|   |||:||:|| |  |       |  |:| : |||  |:::|| :||:|||||  |
Db     141 TKHRPQVYMYAVPYQWIVPFDRIRDGVHAMVAQSVRRTPRSSIDPQVKNFQWGDLIRAVQ 200

Qy     210 EAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGI 269
           |  |||    :||| |  :||| |||| ::||| : || | ||||||||| |:|: :|
Db     201 ETHDRGFEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTVLEIAESLGH 260

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
           || | |:|  || ||||::  |||||| |  |:|| : :| |||:| :|  |:| |  |
Db     261 EAILADITLAELLDADEVLGCTTAGGVWPFVSVDGNPISDGVPGPITQSIIRRYWELNVE 320

Qy     330 PGPLIETIEY 339
               |:  ::|
Db     321 SSSLLTPVQY 330
```

RESULT 6
AAY28430
ID AAY28430 standard; Protein; 310 AA.
XX
AC AAY28430;
XX
DT 15-FEB-2000 (first entry)
XX
DE Corn branched chain amino acid transaminase amino acid sequence.
XX
KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuC; leuD;
KW branched chain amino acid transferase; biosynthetic enzyme; antibody;
KW 3-isopropylmalate dehydratase.
XX
OS Zea mays.

XX
PN WO9921880-A2.
XX
PD 06-MAY-1999.
XX
PF 20-OCT-1998; 98WO-US22081.
XX
PR 28-OCT-1997; 97US-0063423.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Falco SC, Cahoon RE, Hitz WD, Kinney AJ, Rafalski JA;
XX
DR WPI; 2000-022904/02.
DR N-PSDB; AAX89455.
XX
PT Nucleic acid fragments encoding branched chain amino acid biosynthetic
PT enzymes -
XX
PS Claim 6; Fig 5; 102pp; English.
XX
CC AAY28418-Y28431 are fragments of corn, soybean, wheat and rice branched
CC chain amino acid transaminase, amino acid sequences. Sequences
CC AAX89442-X89465 are nucleic acid fragments that encode all or a
CC substantial portion of dihydroxyacid dehydratase, a branched chain amino
CC acid transferase, a leuC or a leuD subunit of 3-isopropylmalate
CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved
CC in biosynthesis and utilization of branched-chain amino acids. The
CC nucleic acid sequences can be used to alter the level of expression of a
CC branched chain amino acid biosynthetic enzyme in a host cell. They can
CC also be used to obtain a nucleic acid fragment encoding a branched chain
CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched
CC chain amino acid biosynthetic enzyme. The encoded proteins may be used
CC to prepare antibodies for detecting the proteins in situ in cells, or in
CC vitro in cell extracts.
XX
SQ Sequence 310 AA;

Query Match 18.3%; Score 322.5; DB 21; Length 310;
Best Local Similarity 29.0%; Pred. No. 1e-23;
Matches 82; Conservative 54; Mismatches 134; Indels 13; Gaps 5;

```
Qy      47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           : |:  | ||   ||:|:||:     |  :    :: | :|:| :|||| | |  :
Db      18 LVWVGDELLPRNSAKVSVFDSVVQGGDAVWEGLRIYDGKVFKLDEHLDRLFDSAKAMAFS 77

Qy     107 AGYSKDELAE-ITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL- 164
           .    ::| : : | |  ::      : : ||:||      |:|  | ::    :|    :
Db      78 NVPTRDWIKDAIFKTLIANGMFNNAHIRLTLTR------GKKVTSGMSPAFNLYGCALIV 131

Qy     165 ---WAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAIL 221
               |  |  :   |  :|    ||   |::|| | :    :  | |     | ||:
Db     132 LAEWKPPVYDNSHG-IKLVTATTRRNSPNSIDPKIHHNNLINNILAKIEGNLAQAEDAIM 190

Qy     222 LDSDNCVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVFELADQMGIEATLRDVTSRE 280
           || |  |:|    |: :|| | : :|  :  ||||||| || :|  :       | ::  |
Db     191 LDKDGFVSETNATNIFMVKKGIVLTPHADYCLPGITRATVMDLVVKENFVLHERRISLSE 250

Qy     281 LYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           : |||:    | | :||:  :||  :|:|: ||:|  |:  :
Db     251 FHAADEVWTTGTMGEITPVVMIDGREIGDGKIGPVTRQIQKAY 293
```

RESULT 7
ABB92207
ID ABB92207 standard; Protein; 490 AA.
XX
AC ABB92207;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 1418.
XX

```
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K,  Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms -
XX
PS   Claim 5; SEQ ID NO 1418; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   490 AA;

  Query Match             18.1%;  Score 318.5;  DB 23;  Length 490;
  Best Local Similarity   28.0%;  Pred. No. 5.1e-23;
  Matches   82;  Conservative   51;  Mismatches  123;  Indels   37;  Gaps    3;

Qy      47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           :||:  | || | ||:|:||:     |  :    :: | ||:| :||||| | |. |  |
Db     232 LAWVGDEILPREMAKVSVFDSVVQGGDSVWEGLRIYKGKIFKLEEHLDRLFDSAKALAFD 291

Qy     107 AGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWA 166
             :::|: |   : :  : | |                                    |
Db     292 NVPAREEVKEAIFRTLITNVLAE----------------------------------WK 316

Qy     167 FPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDN 226
            ||     |   :|    ||   | :|  | :    :   |  |: :  |  ||:|| |
Db     317 -PPVYDNEGGIVLVTATTRRNSPNNLDSKIHHNNLLNNILAKIESNNTNAADAIMLDKDG 375

Qy     227 CVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVFELADQMGIEATLRDVTSRELYDAD 285
           |:|    |: :|| | : :|  :  ||||||  || ||  :       | ::  | : |:
Db     376 YVSETNATNIFMVKKGCVLTPHADYCLPGITRATVMELVVKENFILEERRISLSEFHTAN 435

Qy     286 ELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIE 338
           |:    | | ::|:  :||  :|:|: ||:|  ::: :  | :: |  | | :
Db     436 EVWTTGTMGELSPVVKIDGRVIGDGKVGPVTRTLQNAYKKLTEDSGVPIPTYQ 488


RESULT 8
AAW24255
ID   AAW24255 standard; Protein; 303 AA.
XX
AC   AAW24255;
XX
DT   18-FEB-1998  (first entry)
XX
DE   Pyrobacullum aerophilum branched chain aminotransferase.
XX
KW   Branched chain aminotransferase; IM2/BCA; chiral compound.
XX
OS   Pyrobacullum aerophilum strain IM2.
XX
PN   WO9729187-A1.
XX
PD   14-AUG-1997.
XX
PF   21-JAN-1997;   97WO-US01094.
XX
PR   08-MAY-1996;   96US-0646590.
PR   09-FEB-1996;   96US-0599171.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Swanson RV,  Warren PV;
XX
DR   WPI; 1997-415343/38.
DR   N-PSDB; AAT78779.
XX
PT   New transaminase(s) and aminotransferase(s) derived from host cells
PT   - used for producing enzymes, and hybridisation probes for a cDNA or
PT   genomic library
XX
PS   Claim 19; Fig 8; 95pp; English.
XX
CC   This protein sequence comprises branched chain aminotransferase
CC   IM2/BCA of Pyrobacullum aerophilum IM2, a thermophilic sulphur
CC   archaea that grows optimally at 100 deg C and pH 7.0 in low salt
CC   medium.  The IM2/BCA amino acid sequence was deduced from genomic
CC   DNA (see AAT78779).  Claimed thermostable transaminases and
CC   aminotransferases (AAW24248-57) can be produced from native or
CC   recombinant host cells for use with L- and/or D-amino acids for
CC   production of optically pure chiral compounds used in the
CC   pharmaceutical, agricultural and other industries.  A method is
CC   claimed for transferring an amino group from an amino acid to an
CC   alpha-keto acid using a claimed enzyme.
XX
SQ   Sequence   303 AA;

  Query Match             17.9%;  Score 315;  DB 18;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 5.6e-23;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :     :   :  :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIFEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||       :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
             |   : :   |           ||   : |  ::      |   ||: | :      :
Db     113 RNLEVSLAVIVFP----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
             :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | |||| |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| |||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
             : :
Db     281 AELY 284


RESULT 9
AAW73375
ID   AAW73375 standard; Protein; 282 AA.
XX
AC   AAW73375;
```

```
XX
DT   12-FEB-1999  (first entry)
XX
DE   S. aureus D-alanine transferase protein sequence.
XX
KW   D-alanine transferase; gene therapy; immune response; respiratory tract;
KW   bacterial infection; cardiac disease; gastrointestinal disorder;
KW   central nervous system disorder; eye disease; urinary tract disorder.
XX
OS   Staphylococcus aureus.
XX
PN   EP881285-A1.
XX
PD   02-DEC-1998.
XX
PF   26-MAY-1998;   98EP-0304139.
XX
PR   29-MAY-1997;   97US-0865238.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Warren RL;
XX
DR   WPI; 1999-001391/01.
DR   N-PSDB; AAV08568.
XX
PT   New d-alanine transferase (dat) polypeptide and polynucleotide -
PT   useful as diagnostic reagents and for prevention and treatment of
PT   gastrointestinal, respiratory and CNS bacterial infections
XX
PS   Claim 11; Page 26-27; 31pp; English.
XX
CC   This sequence is the d-alanine transferase (dat) protein of the
CC   invention. Dat polypeptides and polynucleotides are useful for diagnosing
CC   susceptibility to diseases by detecting mutations or polymorphisms of the
CC   dat gene. They can diagnose diseases caused by organisms comprising the
CC   dat gene by detection at the nucleic acid level, and analysing for the
CC   presence or amount of the dat polypeptide in tissue samples. Dat
CC   polypeptides and polynucleotides can be used to screen for agonists and
CC   antagonists which can be used in treatment to enhance or block dat
CC   activity. Gene therapy may also be used to affect endogenous dat
CC   polypeptide production. An immunological response can be induced by
CC   administering the dat polypeptide, to induce an antibody and/or T cell
CC   immune response to protect against disease. This method is especially
CC   useful for preventing bacterial infections (especially S. aureus) from
CC   surgical implants e.g. pacemakers. The dat polypeptide may be used as a
CC   wound treatment to prevent adhesion of bacteria to matrix proteins. Dat
CC   antibodies are also useful for inducing an immune response to immunise
CC   and prevent disease, and for isolating dat clones or purifying the
CC   polypeptide by affinity chromatography. Diseases diagnosed, prevented or
CC   treated include: infections of the upper and lower respiratory tract,
CC   cardiac diseases, gastrointestinal disorders, central nervous system
CC   disorders, eye diseases, kidney and urinary tract disorders, skin
CC   disorders, and bone and joint disorders. Dat polypeptides are useful for
CC   assaying genetic variation, and for mapping genes to chromosomes,
CC   allowing gene inheritance to be studied through linkage analysis.
XX
SQ   Sequence   282 AA;

  Query Match             16.4%;  Score 290;  DB 20;  Length 282;
  Best Local Similarity   26.7%;  Pred. No. 1.6e-20;
  Matches   77;  Conservative   55;  Mismatches  124;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: ||::   |||:|  | |:   |  |    |::| :| : :| :| |  |::: ||
Db       5 FLNGEFVSPSEAKVSYNDRGYVFGDGIYEYIRVYNGKLFTVTEHYERFLRSANEIGLDLN 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           || :|| |:::| | |:|:    : :  |||  :|       ::   :  |   |   :
Db      65 YSVEELIELSRKLVDMNQIETGAIYIQATRGVAERNHSFPTPEVEPAIVAYTKSYDRPYD 124

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGAR 217
             |               | | |  |::: :|          |:: |  :  |
```

```
Db     125 HLEN--------------GVNGV--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAVE 167

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
           :      | ||   |   :||| :   |  |  | ||||  : ::|:   |
Db     168 A--IQHRGETVTEGSSSNAYAIKDGVIYTHPINNYILNGITRIVIKKIAEDYNIPFKEET 225

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
             |   | :|||::  :|:  |||:  |||| : :|: ||:|  ::: |
Db     226 FTVDFLKNADEVIVSSTSAEVTPVIKLDGEPINDGKVGPITRQLQEGF 273


RESULT 10
AAY13559
ID   AAY13559 standard; Protein; 289 AA.
XX
AC   AAY13559;
XX
DT   30-JUL-1999  (first entry)
XX
DE   L. monocytogenes D-amino acid aminotransferase (dat) gene product.
XX
KW   Listeria; vaccine; auxotrophic; attenuated; mutation; dat gene; dal gene;
KW   T-cell response; HIV-1 antigen; HIV; aminotransferase.
XX
OS   Listeria monocytogenes.
XX
PN   WO9925376-A1.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-US24357.
XX
PR   18-NOV-1997;   97US-0972902.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
XX
PI   Frankel FR,  Portnoy DA;
XX
DR   WPI; 1999-337876/28.
DR   N-PSDB; AAX55766.
XX
PT   Attenuated Listeria strain which is auxotrophic for D-alanine
PT   synthesis
XX
PS   Disclosure; Fig 3A-C; 67pp; English.
XX
CC   The invention provides a vaccine comprising an auxotrophic attenuated
CC   strain of Listeria which expresses an antigen, where the strain comprises
CC   a mutation in at least one gene essential for growth of the Listeria
CC   especially that the strain is auxotrophic for D-alanine. The gene is
CC   selected from the Listeria dat and dal genes. The vaccine can be used in
CC   mammals to stimulate T-cell responses to an antigen, especially an HIV-1
CC   antigen. This is useful in creating effective vaccines especially against
CC   HIV. The attenuated Listeria strain can be used as a bacterial vaccine
CC   vector to raise an immunoresponse in e.g. immunocompromised patient or a
CC   pregnant patient. The present sequence represents the L. monocytogenes
CC   D-amino acid aminotransferase (dat) gene product.
XX
SQ   Sequence   289 AA;

  Query Match             16.4%;  Score 289;  DB 20;  Length 289;
  Best Local Similarity   28.9%;  Pred. No. 2.1e-20;
  Matches   83;  Conservative   51;  Mismatches  123;  Indels   30;  Gaps    6;

Qy      50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109
           :    : |:| : | :| |:   |  | |  :::|  |   :|:|||   |:|: |   |
Db       5 VNNHLVEREDATVDIEDRGYQFGDGVYEVVRLYNGKFFTYNEHIDRLYASAAKIDLVIPY 64

Qy     110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPP 169
           ||:|| |: :| |: : :    | | ||||    :              : ||    ||
Db      65 SKEELRELLEKLVAENNINTGNVYLQVTRGVQNPRN------------HVIPD--DFPL 109

Qy     170 AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGART 218
```

```
Db    110 EGVLTAAAREVPRNERQFVEGGTAITEEDVRWLRCDIKSLNLLGNILAKN-KAHQQNALE 168

Qy    219 AILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALPGITRKTVFELADQMGIEATLRDV 276

Db    169 AILHRGEQ-VTECSASNVSIIKDGVLWTHAADNLILNGITRQVIIDVAKKNGIPVKEADF 227

Qy    277 TSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323

Db    228 TLTDLREADEVFISSTTIEITPITHIDGVQVADGKRGPITAQLHQYF 274


RESULT 11
ABB49432
ID   ABB49432 standard; Protein; 289 AA.
XX
AC   ABB49432;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes protein #2136.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW   vitamin B12; bacterial infection; disease.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C,  Frangeul L,  Couve E,  Rusniok C,  Fsihi H,  Dehoux P;
PI   Dussurget O,  Chetouani F,  Nedjari H,  Glaser P,  Kunst F,  Cossart P;
PI   Daniels J,  Goebel W,  Kreft J,  Kuhn M,  Ng E,  Vazquez-Boland JA;
PI   Dominguez-Bernal G,  Garrido-Garcia P,  Tierrez-Martinez A,  Amend A;
PI   Chakraborty T,  Domann E,  Hain T,  Berche P,  Charbit A,  Durant L;
PI   Perez-Diaz J,  Baquero F,  Garcia Del Portillo F,  Gomez-Lopez N;
PI   Maduenio E,  De Pablos B,  Wehland J,  Kaerst U,  Entian K,  Hauf J;
PI   Rose M,  Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides  -
XX
PS   Claim 6; SEQ ID No 2137; 192pp; French.
XX
CC   The present invention relates to the genome sequence of Listeria
CC   monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC   it are useful for selecting probes and primers for detecting genes in L.
CC   monocytogenes and related organisms, and for studying genetic
CC   polymorphisms and other genomes. The present sequence is a protein
CC   encoded by the genome sequence of the present invention. Proteins
CC   expressed from the genome sequence are useful for raising specific
CC   antibodies, identification of L. monocytogenes and related organisms, and
CC   for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC   B12. The genome sequence and proteins encoded by it are also useful for
CC   selecting compounds that regulate gene expression and cell replication,
CC   and modulate L. monocytogenes-related diseases. In addition, the genome
CC   sequence and proteins encoded by it are useful in pharmaceutical and
CC   vaccines compositions for the treatment or prevention of infections by L.
CC   monocytogenes and related organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ   Sequence   289 AA;

  Query Match             16.4%;  Score 289;  DB 23;  Length 289;
  Best Local Similarity   28.9%;  Pred. No. 2.1e-20;
  Matches   83;  Conservative   51;  Mismatches  123;  Indels   30;  Gaps    6;

Qy     50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109

Db      5 VNNHLVEREDATVDIEDRGYQFGDGVYEVVRLYNGKFFTYNEHIDRLYASAAKIDLVIPY 64

Qy    110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPP 169

Db     65 SKEELRELLEKLVAENNINTGNVYLQVTRGVQNPRN------------HVIPD--DFPL 109

Qy    170 AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGART 218

Db    110 EGVLTAAAREVPRNERQFVEGGTAITEEDVRWLRCDIKSLNLLGNILAKN-KAHQQNALE 168

Qy    219 AILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALPGITRKTVFELADQMGIEATLRDV 276

Db    169 AILHRGEQ-VTECSASNVSIIKDGVLWTHAADNLILNGITRQVIIDVAKKNGIPVKEADF 227

Qy    277 TSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323

Db    228 TLTDLREADEVFISSTTIEITPITHIDGVQVADGKRGPITAQLHQYF 274


RESULT 12
AAY13560
ID   AAY13560 standard; Protein; 282 AA.
XX
AC   AAY13560;
XX
DT   30-JUL-1999  (first entry)
XX
DE   S. haemolyticus D-amino acid aminotransferase.
XX
KW   Listeria; vaccine; auxotrophic; attenuated; mutation; dat gene; dal gene;
KW   T-cell response; HIV-1 antigen; HIV; aminotransferase.
XX
OS   Staphylococcus haemolyticus.
XX
PN   WO9925376-A1.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-US24357.
XX
PR   18-NOV-1997;   97US-0972902.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
XX
PI   Frankel FR,  Portnoy DA;
XX
DR   WPI; 1999-337876/28.
XX
PT   Attenuated Listeria strain which is auxotrophic for D-alanine
PT   synthesis
XX
PS   Disclosure; Fig 4A-B; 67pp; English.
XX
CC   The invention provides a vaccine comprising an auxotrophic attenuated
CC   strain of Listeria which expresses an antigen, where the strain comprises
CC   a mutation in at least one gene essential for growth of the Listeria
CC   especially that the strain is auxotrophic for D-alanine. The gene is
CC   selected from the Listeria dat and dal genes. The vaccine can be used in
CC   mammals to stimulate T-cell responses to an antigen, especially an HIV-1
CC   antigen. This is useful in creating effective vaccines especially against
CC   HIV. The attenuated Listeria strain can be used as a bacterial vaccine
CC   vector to raise an immunoresponse in e.g. immunocompromised patient or a
CC   pregnant patient. The present sequence represents a deduced D-amino acid
CC   aminotransferase of S. haemolyticus.
XX
```

```
SQ   Sequence   282 AA;

  Query Match             16.0%;  Score 283;  DB 20;  Length 282;
  Best Local Similarity   25.3%;  Pred. No. 8.2e-20;
  Matches   73;  Conservative   59;  Mismatches  124;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :| ||::   |||:|  | |:   |  |     : | :| : :| :| :  ||:::|| |
Db       5 FINGEFIDQNEAKVSYEDRGYVFGDGIYEYIRAYDGKLFTVTEHFERFIRSASEIQLDLG 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           |: :|| :: :: : :: ::   : :  |||   |       ::   :  :|  |   :
Db      65 YTVEELIDVVRELLKVNNIQNGGIYIQATRGVAPRNHSFPTPEVKPVIMAFAKSYDRPYD 124

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW-----------GDLTAASFEAKDRGAR 217
             |               | |    |::: :|           |:: |  :  |
Db     125 DLEN--------------GINAA--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAGE 167

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
             :      | ||   ||  :||| :   |  |  | ||||| :  :::   |
Db     168 A--IQHRGETVTEGASSNVYAIKDGAIYTHPVNNYILNGITRKVIKWISEDEDIPFKEET 225

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |   | :|||::  :|:  |||:  :||| ||:|: ||:|  ::: |
Db     226 FTVEFLKNADEVIVSSTSAEVTPVVKIDGEQVGDGKVGPVTRQLQEGF 273


RESULT 13
AAP81166
ID   AAP81166 standard; protein; 283 AA.
XX
AC   AAP81166;
XX
DT   05-NOV-1990  (first entry)
XX
DE   D-amino acid transaminase.
XX
KW   D-amino acid transaminase (D-ATA); Bacillus sp; ss.
XX
OS   Bacillus sp.
XX
PN   JP63207387-A.
XX
PD   26-AUG-1988.
XX
PF   24-FEB-1987;   87JP-0039173.
XX
PR   24-FEB-1987;   87JP-0039173.
XX
PA   (DAIL ) DAICEL CHEM IND KK.
XX
DR   WPI; 1988-282536/40.
DR   N-PSDB; AAN81507.
XX
PT   DNA sequence coding for D-amino acid transaminase prodn -
PT   used to increase yield of D-ATA by inserting expression promoter before
PT   gene and inserting into host bacteria.
XX
PS   Disclosure; ; 9pp; Japanese.
XX
CC   This protein is encoded by DNA under control of a high expression
CC   promoter. Upon insertion into an appropriate vector and
CC   transformation of a bacterial host, D-ATA can be produced more
CC   efficiently than by natural D-ATA rpoducing bacteria.
XX
SQ   Sequence   283 AA;

  Query Match             15.8%;  Score 278.5;  DB 9;  Length 283;
  Best Local Similarity   27.6%;  Pred. No. 2.3e-19;
  Matches   83;  Conservative   50;  Mismatches  131;  Indels   37;  Gaps    9;

Qy      45 GGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLR 104
           |   |   : :  || ||   | |:   |  | |  |::| :| : :|:|||   | |:|
Db       2 GYTLW-NDQIVKDEEVKIDKEDRGYQFGDGVYEVVKVYNGEMFTVNEHIDRLYASAEKIR 60

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           :  |:||:  ::  : |  ::|    :  ||||  |          ||  
Db      61 ITIPYTKDKFHQLLHELVEKNELNTGHIYFQVTRGTSPR---------AHQ--------- 102

Qy     165 WAFPP---AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW--GDLTA--------ASFEA 211
             ||       | | |   || :    :      ::: :|   |: :         |  ||
Db     103 --FPENTVKPVIIGYTKENPRPLENLEKGVKATFVEDIRWLRCDIKSLNLLGAVLAKQEA 160

Qy     212 KDRGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGI 269
           ::|   || |  :| | ||   ||  :||| |   |: |  | ||||  |   |::: :
Db     161 HEKGCYEAI-LHRNNTVTEGSSSNVFGIKDGILYTDPANNMILKGITRVVVIACANEINM 219

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
             |: |    |||   :|   :||:  :||: : :|: |  |  :: :|   : :
Db     220 PVKEIPFTTHEALKMDELFVTSTTSEITPVIEIDGKLIRDGKVGEWTRKLQKQFETKIPK 279

Qy     330 P 330
           |
Db     280 P 280


RESULT 14
AAY13562
ID   AAY13562 standard; Protein; 283 AA.
XX
AC   AAY13562;
XX
DT   30-JUL-1999  (first entry)
XX
DE   Bacillus D-amino acid aminotransferase.
XX
KW   Listeria; vaccine; auxotrophic; attenuated; mutation; dat gene; dal gene;
KW   T-cell response; HIV-1 antigen; HIV; aminotransferase.
XX
OS   Bacillus sp.
XX
PN   WO9925376-A1.
XX
PD   27-MAY-1999.
XX
PF   13-NOV-1998;   98WO-US24357.
XX
PR   18-NOV-1997;   97US-0972902.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
XX
PI   Frankel FR,  Portnoy DA;
XX
DR   WPI; 1999-337876/28.
XX
PT   Attenuated Listeria strain which is auxotrophic for D-alanine
PT   synthesis
XX
PS   Disclosure; Fig 4A-B; 67pp; English.
XX
CC   The invention provides a vaccine comprising an auxotrophic attenuated
CC   strain of Listeria which expresses an antigen, where the strain comprises
CC   a mutation in at least one gene essential for growth of the Listeria
CC   especially that the strain is auxotrophic for D-alanine. The gene is
CC   selected from the Listeria dat and dal genes. The vaccine can be used in
CC   mammals to stimulate T-cell responses to an antigen, especially an HIV-1
CC   antigen. This is useful in creating effective vaccines especially against
CC   HIV. The attenuated Listeria strain can be used as a bacterial vaccine
CC   vector to raise an immunoresponse in e.g. immunocompromised patient or a
CC   pregnant patient. The present sequence represents a deduced D-amino acid
CC   aminotransferase of Bacillus species.
XX
SQ   Sequence   283 AA;

  Query Match             15.8%;  Score 278.5;  DB 20;  Length 283;
  Best Local Similarity   27.6%;  Pred. No. 2.3e-19;
```

```
  Matches   83;  Conservative   50;  Mismatches  131;  Indels   37;  Gaps    9;

Qy      45 GGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLR 104
           |   |    : :  || ||   | |:   |  | |  |::| :| : :|:|||   | |:|
Db       2 GYTLW-NDQIVKDEEVKIDKEDRGYQFGDGVYEVVKVYNGEMFTVNEHIDRLYASAEKIR 60

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           :   |:||:  ::  : |  ::|    :   ||||   |          ||
Db      61 ITIPYTKDKFHQLLHELVEKNELNTGHIYFQVTRGTSPR---------AHQ--------- 102

Qy     165 WAFPP---AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW--GDLTA--------ASFEA 211
             ||       | | |   || :    :      ::: :|   |: :         |  ||
Db     103 --FPENTVKPVIIGYTKENPRPLENLEKGVKATFVEDIRWLRCDIKSLNLLGAVLAKQEA 160

Qy     212 KDRGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGI 269
            ::|   || |  :| | ||   || :||| |   |: |  | ||||  |   |::: :
Db     161 HEKGCYEAI-LHRNNTVTEGSSSNVFGIKDGILYTHPANNMILKGITRDVVIACANEINM 219

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
               |: |   |||  :|  :||: :||: : :|: |  |  :: :|   : :
Db     220 PVKEIPFTTHEALKMDELFVTSTTSEITPVIEIDGKLIRDGKVGEWTRKLQKQFETKIPK 279

Qy     330 P 330
           |
Db     280 P 280


RESULT 15
ABB06297
ID   ABB06297 standard; Protein; 288 AA.
XX
AC   ABB06297;
XX
DT   01-JUN-2002  (first entry)
XX
DE   Heat resistant D-amino acid aminotransferase protein.
XX
KW   Heat resistant D-amino acid aminotransferase; D-AAT; EC 2.6.1.21;
KW   enzyme.
XX
OS   Unidentified.
XX
PN   KR99039710-A.
XX
PD   05-JUN-1999.
XX
PF   13-NOV-1997;   97KR-0059891.
XX
PR   13-NOV-1997;   97KR-0059891.
XX
PA   (CHEI-) CHEIL FOODS & CHEM INC.
PA   (KOAD ) KOREA INST SCI & TECHNOLOGY.
XX
PI   Seong MH,  Hong SP,  Kwak MS,  Kim YH,  Suh SH,  Lee JH,  Lee SG;
XX
DR   WPI; 2000-396907/34.
DR   N-PSDB; ABL49611.
XX
PT   Recombinant heat-resistant D-amino acid aminotransferase and process
PT   for preparing D-amino acid using the same  -
XX
PS   Claim 1; Page 6; 9pp; Korean.
XX
CC   The present sequence represents a heat-resistant D-amino acid
CC   aminotransferase (EC 2.6.1.21). The present invention also describes a
CC   process for preparing D-amino acids using the heat resistant D-amino
CC   acid aminotransferase.
XX
SQ   Sequence   288 AA;

  Query Match             14.5%;  Score 256;  DB 21;  Length 288;
  Best Local Similarity   26.0%;  Pred. No. 4.4e-17;
  Matches   73;  Conservative   49;  Mismatches  123;  Indels   36;  Gaps    6;

Qy      51 EGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYS 110
           | ::||  |    : : |    |  | || :: |  | | :|:|||   |: :||   :
Db      10 EKQFLPRHEVTYPMEERGMQFGDGVYEVARIYQGTYFLLEEHIDRLYRSAAAIRLSVPFE 69

Qy     111 KDELAEITKKCVSMSQLRE-SFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPP 169
           || | |  :    |: ::| : : | ||||                    : |   |||
Db      70 KDVLMEKLELLREMNNVKEDAILYLQVTRG-------------------SFPRNHAFPA 109

Qy     170 AEQ--IFGTTAIVPRHVRRAGRNTVDPTIKNYQWG----------DLTAASFEAKDRGAR 217
             :  ::     :|| ::          |: :|                |  || :| |
Db     110 ENRPNLYAYIREMPRKMQEIENGVRTILTKDVRWEYCYIKSLNLLPNVLAKQEAVERQAF 169

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKLAS--PSRNALPGITRKTVFELADQMGIEATLRD 275
           || |  |  : ||   |: :|||| : :   :   | || |  | :   ::||
Db     170 EAI-LHRDGIITEGSSSNIFLVKDGNVYTHPATERILNGIVRMKVKQFCSELGIPLIEEA 228

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLT 316
           :  :: :|||:   :|   : ||  :: : ||:|:|| :|
Db     229 FSINDIAEADEMFLTSTTSSIIPITQVEEQVVGDGKPGEVT 269


Search completed: April 17, 2003, 17:37:23
Job time : 42 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:35:56 ; Search time 19 Seconds
(without alignments)
1715.240 Million cell updates/sec

Title: US-10-067-291-1
Perfect score: 1764
Sequence: 1 MTALSDLGTSNLVAVEPGAI.....RDRFWALMDEPGPLIETIEY 339

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	407.5	23.1	290	2	E69366	branched-chain ami
2	385	21.8	288	2	G64425	branched-chain ami
3	319	18.1	330	2	F69057	branched-chain ami
4	317.5	18.0	311	2	C70463	branched-chain ami
5	304.5	17.3	307	2	A83021	branched-chain ami
6	303.5	17.2	308	1	A34082	branched-chain-ami
7	302.5	17.1	309	2	AD0924	branched-chain ami
8	301	17.1	308	2	AI0474	branched-chain-ami
9	298.5	16.9	309	2	H91216	branched-chain ami
10	298.5	16.9	309	2	A86063	branched-chain ami
11	298	16.9	308	2	B84362	branched-chain ami
12	296.5	16.8	309	1	XNECV	branched-chain-ami
13	290	16.4	282	2	B89960	D-alanine aminotra
14	289	16.4	289	2	AC1277	D-Amino Acid Amino
15	278.5	15.8	282	2	E69829	D-alanine aminotra
16	278.5	15.8	283	2	A31422	D-alanine transami
17	277.5	15.7	319	2	G82374	branched-chain ami
18	273.5	15.5	296	2	G87611	branched-chain ami
19	273	15.5	293	2	AB3555	D-alanine transami
20	264	15.0	289	2	AC1640	D-Amino Acid Amino
21	260	14.7	283	2	C84001	D-alanine aminotra
22	259	14.7	288	2	AD2885	D-alanine aminotra
23	259	14.7	319	2	A97661	D-alanine aminotra
24	253	14.3	304	2	D81445	branched-chain-ami
25	244.5	13.9	285	2	D84197	branched-chain ami
26	244	13.8	288	2	AI3217	D-alanine aminotra
27	240	13.6	287	2	D87465	D-alanine aminotra
28	238.5	13.5	283	2	G95267	probable D-alanine
29	212	12.0	290	2	B97774	hypothetical prote
30	212	12.0	305	2	S75945	hypothetical prote
31	210.5	11.9	273	2	C72328	transaminase B hom
32	210	11.9	290	2	C71701	branched-chain ami
33	205.5	11.6	280	2	E96997	D-amino acid amino
34	204.5	11.6	271	2	G83275	4-amino-4-deoxycho
35	164.5	9.3	410	2	T46034	branched-chain-ami
36	160.5	9.1	295	2	D83661	4-amino-4-deoxycho
37	157	8.9	358	2	D89823	hypothetical prote
38	153	8.7	293	2	C37854	aminodeoxychorisma
39	144	8.2	384	2	T00625	branched-chain ami
40	141	8.0	1421	2	T05892	hypothetical prote
41	139.5	7.9	318	2	T00626	branched-chain ami
42	137	7.8	367	2	C96537	hypothetical prote
43	136	7.7	293	2	AG2683	branched-chain-ami
44	136	7.7	342	2	E97465	brancheD-chain ami
45	132.5	7.5	280	2	D97174	enzyme of ILVE/PAB

ALIGNMENTS

RESULT 1
E69366
branched-chain amino acid aminotransferase (ilvE) homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 18-Jun-1999
C;Accession: E69366
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: E69366
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-290 <KLE>
A;Cross-references: GB:AE001039; GB:AE000782; NID:g2689362; PIDN:AAB90305.1; PID:g264
C;Superfamily: branched-chain-amino-acid aminotransferase

Query Match 23.1%; Score 407.5; DB 2; Length 290;
Best Local Similarity 33.1%; Pred. No. 2.7e-25;
Matches 100; Conservative 60; Mismatches 111; Indels 31; Gaps 7;

```
Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :::||::|  |||:|||| || : |  :     ::| :||| :|:||| | |  : |:
Db       5 YMDGEFVPENEAKVSIFDHGFLYGDGVFEGIRAYNGRVFRLKEHIDRLYDSAKAIDLEIP 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYG-----KRKGE--------KDLSKLTHQ 155
           :|:|  ||  : :  : ||:::   |||| |      || :        |   ||
Db      65 ITKEEFMEIILETLRKNNLRDAYIRPIVTRGIGDLGLDPRKCQNPSIIVITKPWGKLYGD 124

Qy     156 VYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRG 215
           :|             |:  | |||    |||   : : | ||:  : :   |  ||  :|
Db     125 LY------------EK--GLTAITVA-VRRNSFDALPPNIKSLNYLNNILAKIEANAKG 168

Qy     216 ARTAILLDSDNCVAEGPGFNVCIVKDGKLAS-PSRNALPGITRKTVFELADQMGIEATLR 274
             || ||  :  |:|| | |: :||:| : : |: | | ||||: | |: :::||
Db     169 GDEAIFLDRNGYVSEGSGDNIFVVKNGAITTPPTINNLRGITREAVIEIINRLGIPFKET 228

Qy     275 DVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPG-PL 333
           ::   :|| |||:  . ||  : ||  :||  :|:|:|| :|  : : |  | :  | |:
Db     229 NIGLYDLYTADEVFVTGTAAEIAPIVVIDGRKIGDGKPGEITRKLMEEFSKLTESEGVPI 288

Qy     334 IE 335
           |
Db     289 YE 290
```

RESULT 2
G64425

branched-chain amino acid aminotransferase - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: G64425
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: G64425
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-288 <BUL>
A;Cross-references: GB:U67543; GB:L77117; NID:g1591663; PIDN:AAB99010.1; PID:g1591667; T
C;Genetics:
A;Map position: REV939094-938228
C;Superfamily: branched-chain-amino-acid aminotransferase

Query Match 21.8%; Score 385; DB 2; Length 288;
Best Local Similarity 31.9%; Pred. No. 1.8e-23;
Matches 90; Conservative 58; Mismatches 116; Indels 18; Gaps 5;

```
Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: |:::  ::||:|:|| |  : |  :     : | :| | :|:||| | |  | :|
Db       4 YLNGKFVDEKDAKVSVFDHGLLYGDGVFEGIRAYDGVVFMLKEHIDRLYDSAKSLCIDIP 63

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYG-----KRKGEKDLSKLTHQVYIYAIPY 163
           :|:|: ::  : : :: ||:::: | |||| |      ||  |.       ::  |||
Db      64 LTKEEMIDVVLETLRVNNLRDAYIRLVVTRGVGDLGLDPRKCGKP------TIFCIAIP- 116

Qy     164 LWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
           ||     |  ||    |||   : ::| :|:  : :   |  :|   |   | |||
Db     117 ---MPPLLGEDGIRAITV-SVRRLPVDVLNPAVKSLNYLNSVLAKIQANYAGVDEAFLLD 172

Qy     224 SDNCVAEGPGFNVCIVKDGKLASPS--RNALPGITRKTVFELADQMGIEATLRDVTSREL 281
           | || | |: |||:| | :|   :: | ||||  | :|| : |||     :|  :|
Db     173 DKGFVVEGTGDNIFIVKNGVLKTPPVYQSILKGITRDVVIKLAKEEGIEVVEEPLTLHDL 232

Qy     282 YDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           | ||||    ||  : |:  :||  : | : | :|  ::::|
Db     233 YTADELFITGTAAEIVPVFEIDGRVINNKQVGEITKKLKEKF 274
```

RESULT 3
F69057
branched-chain amino-acid aminotransferase - Methanobacterium thermoautotrophicum (strai
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 18-Jun-1999
C;Accession: F69057
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: F69057
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-330 <MTH>
A;Cross-references: GB:AE000905; GB:AE000666; NID:g2622541; PIDN:AAB85907.1; PID:g262254
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1430
A;Start codon: TTG
C;Superfamily: branched-chain-amino-acid aminotransferase

Query Match 18.1%; Score 319; DB 2; Length 330;
Best Local Similarity 30.2%; Pred. No. 4.5e-18;
Matches 93; Conservative 50; Mismatches 121; Indels 44; Gaps 9;

```
Qy      44 AGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWH----------------GNI 86
           | |  |: || :  ||| :            : ::|| |                   |
Db      29 ASGKIWLNGEMVEWEEATV------------HVLSHVVHYGSSVFEGIRCYRNSKGSAI 75

Qy      87 FRLGDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGE 146
           ||| :|: || | |   |:|  |::::: :   : |  : | | ::   | |||    ||
Db      76 FRLREHVKRLFDSAKIYRMDIPYTQEQICDAIVETVRENGLEECYIRPVVFRGY----GE 131

Qy     147 KDLSKLTHQVYIYAIPYLW-AFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQW 201
           :  :   | :   : | |: || :  |  | |    ||   ||:    |   ||
Db     132 MGVHPVNCPVDVAVAAWEWGAYLGAEALEVGVDAGVSTW-RRMAPNTMPNMAKAGGNYLN 190

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           |  |  ||   |   ||:||   ::|| | |: :| :|:: :|  | : | |||| :
Db     191 SQL--AKMEAVRHGYDEAIMLDYHGYISEGSGENIFLVSEGEIYTPPVSSSLLRGITRDS 248

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | ::|   |:      :|   || |||     ||  :||| |:||  :| |  ||:|  :
Db     249 VIKIARTEGVTVHEEPITREMLYIADEAFFTGTAAEITPIRSVDGIEIGAGRRGPVTKLL 308

Qy     320 RDRFWALM 327
           :| |: ::
Db     309 QDEFFRII 316
```

RESULT 4
C70463
branched-chain amino acid aminotransferase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Jun-1999
C;Accession: C70463
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: C70463
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-311 <AQF>
A;Cross-references: GB:AE000762; NID:g2984163; PIDN:AAC07697.1; PID:g2984164; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: ilvE
C;Superfamily: branched-chain-amino-acid aminotransferase

Query Match 18.0%; Score 317.5; DB 2; Length 311;
Best Local Similarity 29.5%; Pred. No. 5.4e-18;
Matches 92; Conservative 46; Mismatches 119; Indels 55; Gaps 7;

```
Qy      43 FAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN-----IFRLGDHLDRL 96
           |:   |: ||: :| |||||||    | :    :  :   |:       |    :| :||
Db       5 FSMDFAFFEGKIVPVEEAKISIMTNSFHYGTAIFEGIRAYWNEEEEQLYILFAKEHYERL 64

Qy      97 LDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQV 156
           |  |  | ::  || :|| |||: :  |::||                          |
Db      65 LTNARCLFMELNYSAEELVEITKEILRKSEIRE------------------------DV 99

Qy     157 YIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGR-----NTVDPTIKNYQWGD-------- 203
           ||  | |        ::   |  :  :: | ||       :    : :::  |
Db     100 YIRPIAYFKDLKLTPKLIDYTPEIAIYLYRFGRYLDTSKGIRAKVSSWRRNDDNSIPSRW 159

Qy     204 --------LTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALP 253
                   |  ||   |  ||||:|   |||| | |: |:|:||  :|| |   |
Db     160 KVAGAYVNSALAKTEALMSGYDEAILLNSQGYVAEGSGENIFIIKNGKAITPSPNEHILE 219

Qy     254 GITRKTVFE-LADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEP 312
           ||||  |   |  :: :|   | :   ||| |||:    ||  |||:  :|   :||||
Db     220 GITRNAVITLLKKELVVEVEERPIARSELYTADEVFLTGTAAEVTPVVEIDNRKIGNGEI 279

Qy     313 GPLTVAIRDRFW 324
           ||:|  ::: ::
Db     280 GPITKQLQEFYF 291
```

RESULT 5
A83021
branched-chain amino acid transferase PA5013 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83021
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83021
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-307 <STO>
A;Cross-references: GB:AE004913; GB:AE004091; NID:g9951286; PIDN:AAG08398.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: ilvE; PA5013
C;Superfamily: branched-chain-amino-acid aminotransferase

Query Match 17.3%; Score 304.5; DB 2; Length 307;
Best Local Similarity 30.0%; Pred. No. 6e-17;
Matches 93; Conservative 47; Mismatches 117; Indels 53; Gaps 10;

```
Qy      46 GVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAH---VWHG----------NIFRLGDH 92
           || | :|| :   :|   :         ||:|: :   |: |           ||||  |
Db       8 GVIWYDGELVQWRDATTHV---------LTHTLHYGMGVFEGVRAYDTPQGTAIFRLQAH 58

Qy      93 LDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTV---TRGYGKR-KGEKD 148
            ||| | |  : :   ||:||: | |:  |  : |  :::   |   : | | |  | |
Db      59 TDRLFDSAHIMNMQIPYSRDEINEATRAAVRENNLESAYIRPMVFYGSEGMGLRASGLK- 117

Qy     149 LSKLTHQVYIYAIPYLWAFPPAEQIF------GTTAIVPRHVR----RAGRNTVDPTIKN 198
                 |::    : |     |:         |::    ||     ||  |
Db     118 -------VHVIIAAWSWGAYMGEEALQQGIKVRTSSFTRHHVNISMTRAKSNGA------ 164

Qy     199 YQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNA-LPGITR 257
             : :   |  ||   ||  |::|| :  |||| | |: |:||| : :|   | | ||||
Db     165 --YINSMLALQEAISGGADEAMMLDPEGYVAEGSGENIFIIKDGVIYTPEVTACLNGITR 222

Qy     258 KTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTV 317
           |:  || : | :   : :|  |:| |||     ||  ||||  :||  :| |  ||:|
Db     223 NTILTLAAEHGFKLVEKRITRDEVYIADEAFFTGTAAEVTPIREVDGRKIGAGRRGPVTE 282

Qy     318 AIRDRFWALM 327
           ::  :: |:
Db     283 KLQKAYFDLV 292
```

RESULT 6
A34082
branched-chain-amino-acid transaminase (EC 2.6.1.42) - Salmonella typhimurium
N;Alternate names: branched-chain-amino-acid aminotransferase; transaminase B
C;Species: Salmonella typhimurium
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C;Accession: A34082; A05077
R;Feild, M.J.; Nguyen, D.C.; Armstrong, F.B.
Biochemistry 28, 5306-5310, 1989
A;Title: Amino acid sequence of Salmonella typhimurium branched-chain amino acid aminotr
A;Reference number: A34082; MUID:89352621; PMID:2669973
A;Accession: A34082
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-308 <FEI>
R;Randall, R.R.; Wallis, M.H.; Young, G.J.; Armstrong, F.B.
Fed. Proc. 38, 325A, 1979
A;Title: N-terminal sequence of branched-chain amino acid aminotransferase.
A;Reference number: A05077
A;Accession: A05077
A;Molecule type: protein
A;Residues: 1-5 <RAN>
C;Function:
A;Description: catalyzes the reversible transamination of alpha-amino groups from leu
A;Pathway: branched-chain amino acid biosynthesis
C;Superfamily: branched-chain-amino-acid aminotransferase
C;Keywords: aminotransferase; branched-chain amino acid biosynthesis; phosphoprotein;
F;159/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 17.2%; Score 303.5; DB 1; Length 308;
Best Local Similarity 31.6%; Pred. No. 7.2e-17;
Matches 91; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

```
Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
           |  || :  |:||: :      :    :     : :    :||  :|: || | |   |
Db       9 WFNGEMVRNEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLRDSAKIYR 68

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           | ||| |  :  :  : |  :::   |  |     |       |  | | | |:
Db      69 FPVSQSIDELMEACRDVIRKNNLTSAYIRPLVFVG-DVGMGVNPPPGYTTDVIIAAFPW- 126

Qy     165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
           |:  || :  |  |:|     ||  ||:    |   ||     |  :  ||:  | :  |
Db     127 GAYLGAEALDQGIDAMVSSW-NRAAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 183

Qy     221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
           || :  ::|| | |:  |||| | :|  : :|||||||  : :|| ::|||   : ::
Db     184 ALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQVLSR 243

Qy     279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWAL 326
             || |||:    ||  :||: |:||  || |  ||:|  |:  |: |
Db     244 ESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGL 291
```

RESULT 7
AD0924
branched-chain amino-acid aminotransferase [imported] - Salmonella enterica subsp. en
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C;Accession: AD0924
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A;Reference number: AB0502; PMID:11677608
A;Accession: AD0924
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD09414.1; PID:g16504532; GSPDB:GN00176
C;Genetics:
A;Gene: STY3654
C;Superfamily: branched-chain-amino-acid aminotransferase

Query Match 17.1%; Score 302.5; DB 2; Length 309;
Best Local Similarity 31.6%; Pred. No. 8.7e-17;
Matches 91; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

```
Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
           |  || :  |:||: :      :    :     : :    :||  :|: || | |   |
Db      10 WFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLRDSAKIYR 69

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           | ||| |  :  :  : |  :::   | |     |       |  | | | |:
Db      70 FPVSQSIDELMEACRDVIRKNNLTSAYIRPLVFVG-DVGMGVNPPPGYTTDVIIAAFPW- 127

Qy     165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
           |:  || :  |  |:|     ||  ||:    |   ||     |  :  ||:  | :  |
Db     128 GAYLGAEALDQGIDAMVSSW-NRAAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 184

Qy     221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
```

```
          || :  ::|| | |:  |||| | :|  : :|||||||  : :|| ::|||   : ::
Db    185 ALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQVLSR 244

Qy    279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWAL 326
            || |||:    ||  :||: |:||  || |  ||:|  |:  |: |
Db    245 ESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGL 292


RESULT 8
AI0474
branched-chain-amino-acid transaminase (EC 2.6.1.42) [imported] - Yersinia pestis (strai
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C;Accession: AI0474
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AI0474
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-308 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC93365.1; PID:g15981811; GSPDB:GN00175
C;Genetics:
A;Gene: ilvE
C;Superfamily: branched-chain-amino-acid aminotransferase
C;Keywords: aminotransferase

  Query Match             17.1%;  Score 301;  DB 2;  Length 308;
  Best Local Similarity   30.2%;  Pred. No. 1.1e-16;
  Matches   92;  Conservative   46;  Mismatches  141;  Indels   26;  Gaps    9;

Qy     49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
           |  || :|  |||: :      :    :     : :    :||  :|: || | |   |
Db      9 WFNGEMVPWAEAKVHVMSHALHYGTSVFEGVRCYESHKGPVVFRHREHMQRLHDSAKIYR 68

Qy    105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           :   | ||| |  :  :  : |  :::   |  |     |          | | | |:
Db     69 MPVSQSVDELMEACRATLRKNNLTSAYIRPLVFIG-DVGMGVNPPEGYNTDVIIAAFPW- 126

Qy    165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
            |:  || :  |  |:|     |   ||:    |   ||    |  :  ||:  | :  |
Db    127 GAYLGAEALEQGIDAMVSSW-NRVAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 183

Qy    221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
            ||    ::|| | |:  |||| | :|  : :|||||||  : :||  ||:|   : ::
Db    184 ALDVHGYLSEGAGENLFEVKDGILFTPPFTSSALPGITRDAIIKLAKDMGLEVREQVLSR 243

Qy    279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF----------WALM 327
            || |||:    ||  :||: |:||  || |: ||:|  |:  |           |  :
Db    244 ESLYLADEVFMSGTAAEITPVRSVDGIQVGIGKRGPVTAKIQQAFFGLFTGETEDKWGWL 303

Qy    328 DEPGP 332
          |:  |
Db    304 DQVNP 308


RESULT 9
H91216
branched-chain amino-acid aminotransferase [imported] - Escherichia coli (strain O157:H7
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: H91216
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H91216
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB38127.1; PID:g13364179; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4704
C;Superfamily: branched-chain-amino-acid aminotransferase

  Query Match             16.9%;  Score 298.5;  DB 2;  Length 309;
  Best Local Similarity   30.9%;  Pred. No. 1.8e-16;
  Matches   89;  Conservative   49;  Mismatches  135;  Indels   15;  Gaps    8;

Qy     49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
           |  || :  |:||: :      :    :     : :    :||  :|: || | |   |
Db     10 WFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLHDSAKIYR 69

Qy    105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
            | ||| |  :  :  : |  :::   :  |     |     :  :  | | | |:
Db     70 FPVSQSIDELMEACRDVIRKNNLTSAYIRPLIFVG-DVGMGVNPPAGYSTDVIIAAFPW- 127

Qy    165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
            |:  || :  |  |:|     ||  ||:    |   ||    |  :  ||:  | :  |
Db    128 GAYLGAEALEQGIDAMVSSW-NRAAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 184

Qy    221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
            || :  ::|| | |:  |||| | :|  : :|||||||  : :||  ::|||   : ::
Db    185 ALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQVLSR 244

Qy    279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWAL 326
            || |||:    ||  :||: |:||  || |  ||:|  |:  |: |
Db    245 ESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGL 292


RESULT 10
A86063
branched-chain amino-acid aminotransferase [imported] - Escherichia coli (strain O157
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: A86063
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86063
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <STO>
A;Cross-references: GB:AE005174; NID:g12518628; PIDN:AAG58965.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ilvE
C;Superfamily: branched-chain-amino-acid aminotransferase

  Query Match             16.9%;  Score 298.5;  DB 2;  Length 309;
  Best Local Similarity   30.9%;  Pred. No. 1.8e-16;
  Matches   89;  Conservative   49;  Mismatches  135;  Indels   15;  Gaps    8;

Qy     49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
           |  || :  |:||: :      :    :     : :    :||  :|: || | |   |
Db     10 WFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLHDSAKIYR 69

Qy    105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
            | ||| |  :  :  : |  :::   :  |     |     :  :  | | | |:
Db     70 FPVSQSIDELMEACRDVIRKNNLTSAYIRPLIFVG-DVGMGVNPPAGYSTDVIIAAFPW- 127

Qy    165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
            |:  || :  |  |:|     ||  ||:    |   ||    |  :  ||:  | :  |
Db    128 GAYLGAEALEQGIDAMVSSW-NRAAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 184

Qy    221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
            || :  ::|| | |:  |||| | :|  : :|||||||  : :||  ::|||   : ::
Db    185 ALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQVLSR 244
```

```
Qy    279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWAL 326
          || |||:    ||  :||: |:||  || |  ||:|  |:  |: |
Db    245 ESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGL 292


RESULT 11
B84362
branched-chain amino acid aminotransferase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: B84362
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: B84362
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-308 <STO>
A;Cross-references: GB:AE004437; NID:g10581537; PIDN:AAG20262.1; GSPDB:GN00138
C;Genetics:
A;Gene: ilvE2
C;Superfamily: branched-chain-amino-acid aminotransferase

  Query Match             16.9%;  Score 298;  DB 2;  Length 308;
  Best Local Similarity   27.7%;  Pred. No. 2e-16;
  Matches   93;  Conservative   56;  Mismatches  127;  Indels   60;  Gaps   12;

Qy     32 YSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IF 87
          :|: ::||         |::||::  |:|:: :      :    :    |:  :    ||
Db      4 FSEMDVDT--------IWMDGEFVDWEDAQVHVLTHALHYGSGIFEGVRVYDTDNGPAIF 55

Qy     88 RLGDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGY---GKRK 144
          |  ||| |  : |    ||  :: ||| : | : :    |   ::      ||   |
Db     56 RWDDHLQRFYESAKPYDLDIEFTPDELTDATVELIQRQDLGSCYIRPLAYYGYESLGVSP 115

Qy    145 GEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQ---- 200
          |:         | | | |  |         |  |:          | ||  :  ::
Db    116 GD-----CPTDVAIAAWP--WG-----AYLGDDAL---------ENGVDVMVSTWRKHAS 154

Qy    201 -----------WGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP- 247
                    : :   |  ||:  |   ||:|: :  |||||| |: :|:||:| :|
Db    155 SQIPTNAKTTGLYVNSMLAGEEARRNGFTEAIVLNKEGNVAEGPGENIFLVRDGELFTPG 214

Qy    248 -SRNALPGITRKTVFELADQMGIEATLRD---VTSRELYDADELMAVTTAGGVTPINSLD 303
           | : | |||| ||  ||:: |   |: |   ::  ||: |||    :|  ||||  :|
Db    215 LSESILDGITRDTVITLAEERGY--TVHDDVSISRGELHTADELFFTGSAAEVTPIKQVD 272

Qy    304 GEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339
            :|:|  ||:|  ::  |: |: | |   |   |
Db    273 NVEIGSGTRGPVTEELQTAFFDLV-EAGDREEWFHY 307


RESULT 12
XNECV
branched-chain-amino-acid transaminase (EC 2.6.1.42) [validated] - Escherichia coli (str
N;Alternate names: branched-chain-amino-acid aminotransferase; transaminase B
C;Species: Escherichia coli
C;Date: 30-Nov-1980 #sequence_revision 10-Oct-1997 #text_change 01-Mar-2002
C;Accession: E65180; C26570; A91996; A93693; S30668; A57546; A00602
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: E65180
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-309 <BLAT>
A;Cross-references: GB:AE000453; GB:U00096; NID:g2367276; PIDN:AAC77490.1; PID:g1790205;
A;Experimental source: strain K-12, substrain MG1655
R;Lawther, R.P.; Wek, R.C.; Lopes, J.M.; Pereira, R.; Taillon, B.E.; Hatfield, G.W.
Nucleic Acids Res. 15, 2137-2155, 1987
A;Title: The complete nucleotide sequence of the ilvGMEDA operon of Escherichia coli
A;Reference number: A26570; MUID:87174741; PMID:3550695
A;Accession: C26570
A;Molecule type: DNA
A;Residues: 1-150,'A',152-309 <LA1>
A;Cross-references: GB:X04890; NID:g288528; PIDN:CAA28575.1; PID:g288531
R;Kuramitsu, S.; Ogawa, T.; Ogawa, H.; Kagamiyama, H.
J. Biochem. 97, 993-999, 1985
A;Title: Branched-chain amino acid aminotransferase of Escherichia coli: nucleotide s
A;Reference number: A91996; MUID:85289113; PMID:3897211
A;Accession: A91996
A;Molecule type: DNA
A;Residues: 1-150,'A',152-309 <KUR>
A;Cross-references: GB:X02413; NID:g41787; PIDN:CAA26262.1; PID:g41789
A;Experimental source: strain K12
R;Lawther, R.P.; Nichols, B.; Zurawski, G.; Hatfield, G.W.
Nucleic Acids Res. 7, 2289-2301, 1979
A;Title: The nucleotide sequence preceding and including the beginning of the ilvE ge
A;Reference number: A93693; MUID:80101081; PMID:392469
A;Accession: A93693
A;Molecule type: DNA
A;Residues: 1-57,'V',59-61,'R',63-81 <LA2>
A;Experimental source: strain K12
R;Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A;Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.
A;Reference number: S30660; MUID:92358234; PMID:1379743
A;Accession: S30668
A;Molecule type: DNA
A;Residues: 1-309 <DAN>
A;Cross-references: EMBL:M87049; NID:g836656; PIDN:AAA67573.1; PID:g148177
R;Inoue, K.; Kuramitsu, S.; Aki, K.; Watanabe, Y.; Takagi, T.; Nishigai, M.; Ikai, A.
J. Biochem. 104, 777-784, 1988
A;Title: Branched-chain amino acid aminotransferase of Escherichia coli: overproducti
A;Reference number: A57546; MUID:89174510; PMID:3069843
A;Accession: A57546
A;Molecule type: protein
A;Residues: 2-31;150,'A',152-165;308-309 <INO>
A;Note: identification of N6-pyridoxal phosphate lysine
C;Genetics:
A;Gene: ilvE
A;Map position: 85 min
C;Function:
A;Description: EC 2.6.1.42 [validated, MUID:85289113]; catalyzes the reversible trans
A;Pathway: branched-chain amino acid biosynthesis
C;Superfamily: branched-chain-amino-acid aminotransferase
C;Keywords: aminotransferase; branched-chain amino acid biosynthesis; phosphoprotein;
F;2-309/Product: branched-chain-amino-acid aminotransferase #status experimental <MAT
F;160/Binding site: pyridoxal phosphate (Lys) (covalent) #status experimental

  Query Match             16.8%;  Score 296.5;  DB 1;  Length 309;
  Best Local Similarity   30.9%;  Pred. No. 2.7e-16;
  Matches   89;  Conservative   49;  Mismatches  135;  Indels   15;  Gaps    8;

Qy     49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
          |  || :  |:||: :      :    :      : :    :||  :|: || | |   |
Db     10 WFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLHDSAKIYR 69

Qy    105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
          | ||| |  :  :  : |  :::   :  |     |    :  :  | | |.|:
Db     70 FPVSQSIDELMEACRDVIRKNNLTSAYIRPLIFVG-DVGMGVNPPAGYSTDVIIAAFPW- 127

Qy    165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
          |:  || :  |  |:|     ||  ||:    |   ||    |  :  ||:  | :  |
Db    128 GAYLGAEALEQGIDAMVSSW-NRARPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 184

Qy    221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
          || :  ::|| | |:  |||| | :|  : :|||||||  : :|| ::|||   : ::
Db    185 ALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQVLSR 244
```

```
Qy    279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWAL 326
          || |||:    ||  :||: |:||  || |  ||:|  |:  |: |
Db    245 ESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGL 292


RESULT 13
B89960
D-alanine aminotransferase [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 01-Mar-2002
C;Accession: B89960
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: B89960
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-282 <KUR>
A;Cross-references: GB:BA000018; PID:g13701546; PIDN:BAB42839.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1571
C;Superfamily: branched-chain-amino-acid aminotransferase

  Query Match             16.4%;  Score 290;  DB 2;  Length 282;
  Best Local Similarity   26.7%;  Pred. No. 7.8e-16;
  Matches   77;  Conservative   55;  Mismatches  124;  Indels   32;  Gaps    6;

Qy     49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
          :: ||::   |||:|  | |:   |  |    |::| :| : :| :| |  |::: ||
Db      5 FLNGEFVSPSEAKVSYNDRGYVFGDGIYEYIRVYNGKLFTVTEHYERFLRSANEIGLDLN 64

Qy    109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
          || :|| |:::| | |:|:    : :  |||  :|       ::   :  |   |   :
Db     65 YSVEELIELSRKLVDMNQIETGAIYIQATRGVAERNHSFPTPEVEPAIVAYTKSYDRPYD 124

Qy    169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW-----------GDLTAASFEAKDRGAR 217
          |          | | |  |::: :|           |:: |  :  |
Db    125 HLEN--------------GVNGV--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAVE 167

Qy    218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
             :      | ||   |   :||| :   |  |  | ||||  : ::|:   |
Db    168 A--IQHRGETVTEGSSSNAYAIKDGVIYTHPINNYILNGITRIVIKKIAEDYNIPFKEET 225

Qy    276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
          |   | :|||::  :|:  |||:  |||| : :|: ||:|  ::: |
Db    226 FTVDFLKNADEVIVSSTSAEVTPVIKLDGEPINDGKVGPITRQLQEGF 273


RESULT 14
AC1277
D-Amino Acid Aminotransferase [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AC1277
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AC1277
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-289 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99697.1; PID:g16411055; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: daaA
C;Superfamily: branched-chain-amino-acid aminotransferase

  Query Match             16.4%;  Score 289;  DB 2;  Length 289;
  Best Local Similarity   28.9%;  Pred. No. 9.7e-16;
  Matches   83;  Conservative   51;  Mismatches  123;  Indels   30;  Gaps    6;

Qy     50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109
          :    :  |:| : | | |:   |  | |  :::|  |   :|:|||   |:|: |   |
Db      5 VNNHLVEREDATVDIEDRGYQFGDGVYEVVRLYNGKFFTYNEHIDRLYASAAKIDLVIPY 64

Qy    110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPP 169
          ||:|| |: :| |: : :    | | ||||    :               : ||    ||
Db     65 SKEELRELLEKLVAENNINTGNVYLQVTRGVQNPRN------------HVIPD--DFPL 109

Qy    170 AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW-----------GDLTAASFEAKDRGART 218
          :       |||: |:         | :: :|           |:: | : :|  : |
Db    110 EGVLTAAAREVPRNERQFVEGGTAITEEDVRWLRCDIKSLNLLGNILAKN-KAHQQNALE 168

Qy    219 AILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALPGITRKTVFELADQMGIEATLRDV 276
          |||   :  | |    || |:||| | : : :   | ||||: : ::| : ||     |
Db    169 AILHRGEQ-VTECSASNVSIIKDGVLWTHAADNLILNGITRQVIIDVAKKNGIPVKEADF 227

Qy    277 TSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
          |  :| :|||:   :|   :|||  :||  | :|: ||:|  :   |
Db    228 TLTDLREADEVFISSTTIEITPITHIDGVQVADGKRGPITAQLHQYF 274


RESULT 15
E69829
D-alanine aminotransferase homolog yheM - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: E69829
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69829
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-282 <KUN>
A;Cross-references: GB:Z99109; GB:AL009126; NID:g2633260; PIDN:CAB12806.1; PID:g26333
A;Experimental source: strain 168
C;Genetics:
A;Gene: yheM
C;Superfamily: branched-chain-amino-acid aminotransferase

  Query Match             15.8%;  Score 278.5;  DB 2;  Length 282;
  Best Local Similarity   29.1%;  Pred. No. 6.6e-15;
  Matches   82;  Conservative   48;  Mismatches  129;  Indels   23;  Gaps    6;

Qy     50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109
          : |  :   || | : | |:   |  | |  |: | :| | :| :|    |::: :   :
Db      5 VNGRLIGRSEASIDLEDRGYQFGDGIYEVIRVYKGVLFGLREHAERFFRSAAEIGISLPF 64

Qy    110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAF-- 167
          | ::|    :| |  : : |  | :  |||   ||          ||      |   |:
Db     65 SIEDLEWDLQKLVQENAVSEGAVYIQTTRGVAPRK---------HQYEAGLEPQTTAYTF 115

Qy    168 ----PPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
```

```
              |  || :|  ||    :|       |    |  :  :|     | : ||  ||||
Db    116 TVKKPEQEQAYGVAAITDEDLRWL---RCDIKSLNLLYNVMTKQ--RAYEAGAFEAILL- 169

Qy    224 SDNCVAEGPGFNVCIVKDGKLAS--PSRNALPGITRKTVFELADQMGIEATLRDVTSREL 281
            |  | ||   ||  | :| : :   :|  | ||||  :  | :: ||:     |:  ||
Db    170 RDGVVTEGTSSNVYAVINGTVRTHPANRLILNGITRMNILGLIEKNGIKLDETPVSEEEL 229

Qy    282 YDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
          |:|:   :|   : |: :|||:::|:|:|||:|  ::  |
Db    230 KQAEEIFISSTTAEIIPVVTLDGQSIGSGKPGPVTKQLQAAF 271
```

Search completed: April 17, 2003, 17:39:55
Job time : 21 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:29:31 ; Search time 25 Seconds
(without alignments)
562.419 Million cell updates/sec

Title: US-10-067-291-1
Perfect score: 1764
Sequence: 1 MTALSDLGTSNLVAVEPGAI.....RDRFWALMDEPGPLIETIEY 339

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	407.5	23.1	290	1	ILVE_ARCFU	O29329 archaeoglob
2	385	21.8	288	1	ILVE_METJA	Q58414 methanococc
3	319	18.1	306	1	ILVE_METTH	O27481 methanobact
4	317.5	18.0	311	1	ILVE_AQUAE	O67733 aquifex aeo
5	309.5	17.5	283	1	DAAA_BACLI	P54692 bacillus li
6	304.5	17.3	307	1	ILVE_PSEAE	O86428 pseudomonas
7	302.5	17.1	308	1	ILVE_SALTY	P15168 salmonella
8	298.5	16.9	308	1	ILVE_ECOLI	P00510 escherichia
9	289	16.4	289	1	DAAA_LISMO	O85046 listeria mo
10	283	16.0	282	1	DAAA_STAHA	P54694 staphylococ
11	278.5	15.8	282	1	DAAA_BACSP	P19938 bacillus sp
12	278.5	15.8	282	1	DAAA_BACSU	O07597 bacillus su
13	264	15.0	289	1	DAAA_LISIN	Q92b90 listeria in
14	250	14.2	283	1	DAAA_BACSH	P54693 bacillus sp
15	212	12.0	290	1	ILVE_RICCN	Q92i26 rickettsia
16	212	12.0	305	1	ILVE_SYNY3	P54691 synechocyst
17	210.5	11.9	273	1	ILVE_THEMA	P74921 thermotoga
18	210	11.9	290	1	ILVE_RICPR	O05970 rickettsia
19	153	8.7	293	1	PABC_BACSU	P28821 bacillus su
20	120.5	6.8	368	1	ILVE_MYCTU	Q10399 mycobacteri
21	119.5	6.8	356	1	YBGE_BACSU	O31461 bacillus su
22	116	6.6	384	1	BCAT_HUMAN	P54687 homo sapien
23	114.5	6.5	267	1	PABC_VIBCH	Q9kqi0 vibrio chol
24	113.5	6.4	368	1	ILVE_MYCLE	O32954 mycobacteri
25	110.5	6.3	386	1	BCAT_MOUSE	P24288 mus musculu
26	109.5	6.2	269	1	PABC_ECOLI	P28305 escherichia
27	109.5	6.2	271	1	PABC_VIBHA	Q56693 vibrio harv
28	102.5	5.8	362	1	ILVE_BACSU	P39576 bacillus su
29	100.5	5.7	362	1	ILVE_STRCO	O86505 streptomyce
30	100.5	5.7	411	1	BCAT_RAT	P54690 rattus norv
31	100	5.7	874	1	SYA_PASMU	P57933 pasteurella
32	99.5	5.6	607	1	DNAK_STRPN	P95829 streptococc
33	99.5	5.6	1473	1	TOP2_ARATH	P30182 arabidopsis
34	99	5.6	638	1	DNAK_PRELO	Q93gf1 prevotella
35	98.5	5.6	807	1	AFSK_STRGR	P54742 streptomyce
36	98	5.6	385	1	ACKA_MYCTU	P96255 mycobacteri
37	96.5	5.5	610	1	DNAK_BACSH	O69268 bacillus sp
38	95.5	5.4	393	1	BCAM_RAT	O35854 rattus norv
39	95	5.4	379	1	Y829_BRUME	Q8yhh1 brucella me
40	95	5.4	1076	1	CARB_HALER	Q8rss3 halomonas e
41	92	5.2	1081	1	CARB_RALSO	Q8xz83 ralstonia s
42	91.5	5.2	343	1	ILVE_HAEIN	P54689 haemophilus
43	91.5	5.2	1014	1	BGAL_BACHD	Q9k9c6 bacillus ha
44	91	5.2	721	1	BBS2_HUMAN	Q9bxc9 homo sapien
45	91	5.2	3038	1	TRIO_HUMAN	O75962 homo sapien

ALIGNMENTS

RESULT 1
ILVE_ARCFU
ID ILVE_ARCFU STANDARD; PRT; 290 AA.
AC O29329;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)
DE (Transaminase B) (BCAT).
GN ILVE OR AF0933.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
CC -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC oxopentanoate + L-glutamate.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: Valine and isoleucine biosynthesis.
CC -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC ----------

CC ----------
DR EMBL; AE001039; AAB90305.1; -.
DR HSSP; P19938; 1A0G.
DR TIGR; AF0933; -.
DR InterPro; IPR001544; Aminotran_4.
DR Pfam; PF01063; aminotran_4; 1.
DR ProDom; PD001961; Aminotran_4; 1.
DR TIGRFAMs; TIGR01122; ilvE_I; 1.
DR PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.

```
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     150    150       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   290 AA;  32366 MW;  F096FF8CFC3BA273 CRC64;

  Query Match             23.1%;  Score 407.5;  DB 1;  Length 290;
  Best Local Similarity   33.1%;  Pred. No. 2.2e-25;
  Matches  100;  Conservative   60;  Mismatches  111;  Indels   31;  Gaps    7;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :::||::|  |||:|||| || : |  :     ::| :||| :|:||| | |  : |:
Db       5 YMDGEFVPENEAKVSIFDHGFLYGDGVFEGIRAYNGRVFRLKEHIDRLYDSAKAIDLEIP 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYG-----KRKGE--------KDLSKLTHQ 155
           :|:|  ||  : :  : ||:::   |||| |      || :        |   ||
Db      65 ITKEEFMEIILETLRKNNLRDAYIRPIVTRGIGDLGLDPRKCQNPSIIVITKPWGKLYGD 124

Qy     156 VYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRG 215
           :|           |: | |||    |||   : : | ||:  : :   |  ||  :|
Db     125 LY------------EK--GLTAITVA-VRRNSFDALPPNIKSLNYLNNILAKIEANAKG 168

Qy     216 ARTAILLDSDNCVAEGPGFNVCIVKDGKLAS-PSRNALPGITRKTVFELADQMGIEATLR 274
           || || :  |:|| | |: :||:| : : |: | | ||||: | |: :::||
Db     169 GDEAIFLDRNGYVSEGSGDNIFVVKNGAITTPPTINNLRGITREAVIEIINRLGIPFKET 228

Qy     275 DVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPG-PL 333
           ::   :|| |||:    ||  : ||  :|| :|:|:|| :|  : : |  | :  | |:
Db     229 NIGLYDLYTADEVFVTGTAAEIAPIVVIDGRKIGDGKPGEITRKLMEEFSKLTESEGVPI 288

Qy     334 IE 335
           |
Db     289 YE 290


RESULT 2
ILVE_METJA
ID   ILVE_METJA     STANDARD;      PRT;   288 AA.
AC   Q58414;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)
DE   (Transaminase B) (BCAT).
GN   ILVE OR MJ1008.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC       oxopentanoate + L-glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- PATHWAY: Valine and isoleucine biosynthesis.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67543; AAB99010.1; -.
DR   HSSP; P00510; 1A3G.
DR   TIGR; MJ1008; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     146    146       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   288 AA;  32096 MW;  033DB417DFD89CBC CRC64;

  Query Match             21.8%;  Score 385;  DB 1;  Length 288;
  Best Local Similarity   31.9%;  Pred. No. 1.4e-23;
  Matches   90;  Conservative   58;  Mismatches  116;  Indels   18;  Gaps    5;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: |:::  ::||:|:|| |  : |  :     : | :| | :|:||| | |  | :|
Db       4 YLNGKFVDEKDAKVSVFDHGLLYGDGVFEGIRAYDGVVFMLKEHIDRLYDSAKSLCIDIP 63

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYG-----KRKGEKDLSKLTHQVYIYAIPY 163
           :|:|: ::  : : :: ||:::: | |||| |      ||  |        ::  |||
Db      64 LTKEEMIDVVLETLRVNNLRDAYIRLVVTRGVGDLGLDPRKCGKP------TIFCIAIP- 116

Qy     164 LWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
               ||     |  ||    |||   : ::| :|:  : :   |  :|   |   | |||
Db     117 ---MPPLLGEDGIRAITV-SVRRLPVDVLNPAVKSLNYLNSVLAKIQANYAGVDEAFLLD 172

Qy     224 SDNCVAEGPGFNVCIVKDGKLASPS--RNALPGITRKTVFELADQMGIEATLRDVTSREL 281
           | || | |: |||:| | :|   :: | ||||  | :|| : |||     :|  :|
Db     173 DKGFVVEGTGDNIFIVKNGVLKTPPVYQSILKGITRDVVIKLAKEEGIEVVEEPLTLHDL 232

Qy     282 YDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           | ||||    ||  : |:  :||  : | : | :|  ::::|
Db     233 YTADELFITGTAAEIVPVFEIDGRVINNKQVGEITKKLKEKF 274


RESULT 3
ILVE_METTH
ID   ILVE_METTH     STANDARD;      PRT;   306 AA.
AC   O27481;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)
DE   (Transaminase B) (BCAT).
GN   ILVE OR MTH1430.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC   Methanobacteriaceae; Methanothermobacter.
OX   NCBI_TaxID=187420;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Delta H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
```

```
CC   -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC       oxopentanoate + L-glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- PATHWAY: Valine and isoleucine biosynthesis.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000905; AAB85907.1; ALT_INIT.
DR   HSSP; P00510; 1A3G.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     159    159       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   306 AA;  34205 MW;  725611CAC4796F98 CRC64;

  Query Match             18.1%;  Score 319;  DB 1;  Length 306;
  Best Local Similarity   30.2%;  Pred. No. 2.7e-18;
  Matches   93;  Conservative   50;  Mismatches  121;  Indels   44;  Gaps    9;

Qy      44 AGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWH----------------GNI 86
           | |  |: || :  ||| :             : ::|| |                   |
Db       5 ASGKIWLNGEMVEWEEATV------------HVLSHVVHYGSSVFEGIRCYRNSKGSAI 51

Qy      87 FRLGDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGE 146
           ||| :|: || | |   |:|  |::::: :   : |  : | | ::   | |||    ||
Db      52 FRLREHVKRLFDSAKIYRMDIPYTQEQICDAIVETVRENGLEECYIRPVVFRGY----GE 107

Qy     147 KDLSKLTHQVYIYAIPYLW-AFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQW 201
             :  :   | :    : | |:  || :  |  | |    ||   ||:    |   ||
Db     108 MGVHPVNCPVDVAVAAWEWGAYLGAEALEVGVDAGVSTW-RRMAPNTMPNMAKAGGNYLN 166

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           |  |  ||   |   ||:||    ::|| | |: :| :|:: :|  | : | |||| :
Db     167 SQL--AKMEAVRHGYDEAIMLDYHGYISEGSGENIFLVSEGEIYTPPVSSSLLRGITRDS 224

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | ::|   |:      :|   || |||     ||  :||| |:||  :| |  ||:|  :
Db     225 VIKIARTEGVTVHEEPITREMLYIADEAFFTGTAAEITPIRSVDGIEIGAGRRGPVTKLL 284

Qy     320 RDRFWALM 327
           :| |: ::
Db     285 QDEFFRII 292


RESULT 4
ILVE_AQUAE
ID   ILVE_AQUAE     STANDARD;      PRT;   311 AA.
AC   O67733;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable branched-chain amino acid aminotransferase (EC 2.6.1.42)
DE   (BCAT).
GN   ILVE OR AQ_1893.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC   Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC       oxopentanoate + L-glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Valine and isoleucine biosynthesis.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000762; AAC07697.1; -.
DR   HSSP; P00510; 1A3G.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     160    160       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   311 AA;  35763 MW;  0C02F2520947D647 CRC64;

  Query Match             18.0%;  Score 317.5;  DB 1;  Length 311;
  Best Local Similarity   29.5%;  Pred. No. 3.6e-18;
  Matches   92;  Conservative   46;  Mismatches  119;  Indels   55;  Gaps    7;

Qy      43 FAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN-----IFRLGDHLDRL 96
           |:   |: ||: :| |||||||    | :    :  :   |:       |    :| :||
Db       5 FSMDFAFFEGKIVPVEEAKISIMTNSFHYGTAIFEGIRAYWNEEEEQLYILFAKEHYERL 64

Qy      97 LDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQV 156
           |  |  | ::  || :|| ||||: :  |::||                          |
Db      65 LTNARCLFMELNYSAEELVEITKEILRKSEIRE-----------------------DV 99

Qy     157 YIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGR-----NTVDPTIKNYQWGD-------- 203
           ||  | |         ::   |  :  :: | ||       :   : :::  |
Db     100 YIRPIAYFKDLKLTPKLIDYTPEIAIYLYRFGRYLDTSKGIRAKVSSWRRNDDNSIPSRW 159

Qy     204 --------LTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALP 253
                   |  ||   |  ||||:|   |||| | |: |:|:||  :|| |   |
Db     160 KVAGAYVNSALAKTEALMSGYDEAILLNSQGYVAEGSGENIFIIKNGKAITPSPNEHILE 219

Qy     254 GITRKTVFE-LADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEP 312
           ||||  |   |  :: :|   | :   ||| |||:    ||  |||:  :|   :||||
Db     220 GITRNAVITLLKKELVVEVEERPIARSELYTADEVFLTGTAAEVTPVVEIDNRKIGNGEI 279

Qy     313 GPLTVAIRDRFW 324
           ||:|  ::: ::
Db     280 GPITKQLQEFYF 291


RESULT 5
DAAA_BACLI
ID   DAAA_BACLI     STANDARD;      PRT;   283 AA.
AC   P54692;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
DE   aminotransferase) (D-amino acid aminotransferase) (D-amino acid
```

```
DE   transaminase) (DAAT).
GN   DAT.
OS   Bacillus licheniformis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1402;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10716;
RX   MEDLINE=97157074; PubMed=9003455;
RA   Taylor P.P., Fotheringham I.G.;
RT   "Nucleotide sequence of the Bacillus licheniformis ATCC 10716 dat
RT   gene and comparison of the predicted amino acid sequence with those
RT   of other bacterial species.";
RL   Biochim. Biophys. Acta 1350:38-40(1997).
CC   -!- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,
CC       GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-
CC       glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U26947; AAB50428.1; -.
DR   HSSP; P19938; 1DAA.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Pyridoxal phosphate.
FT   BINDING     144    144       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   283 AA;  31396 MW;  2CA8FA604FEDE9D7 CRC64;

  Query Match             17.5%;  Score 309.5;  DB 1;  Length 283;
  Best Local Similarity   30.7%;  Pred. No. 1.4e-17;
  Matches   85;  Conservative   49;  Mismatches  112;  Indels   31;  Gaps    7;

Qy      52 GEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSK 111
           |  :   |  : | | |:   |  | |  :::| :| | :|: ||   |::: :|  :|:
Db       7 GRLMERSECAVDIEDRGYQFGDGVYEVIRIYNGILFTLDEHIARLYKSAAEIGIDLSFSE 66

Qy     112 DELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAE 171
            ||    |: | ::| |:  : | ||||   || :   : || ||  |  |      |  |
Db      67 AELKSQLKELVDINQRRDGGLYLQVTRGKAPRKHQYG-AGLTPQVTAYTFPI--QKPEKE 123

Qy     172 QIFGTTAIVPRHVRRAGRNTVDPTIKNYQW--GDLTAASF--------EAKDRGARTAIL 221
           |  | :||               |  : :|   |: : :         :|::  |  |||
Db     124 QQNGVSAI--------------TADDMRWLRCDIKSLNLLYNVMIKQKAQEASAFEAIL 168

Qy     222 LDSDNCVAEGPGFNVCIVKDGKLAS--PSRNALPGITRKTVFELADQMGIEATLRDVTSR 279
           :  |  | ||   || : |   :  :   :   | ||||  | :| :: |:    : ||
Db     169 I-RDGLVTEGTSSNVYVAKQNVIYTHPVTTLILNGITRMKVLQLCEENGLNYEEKAVTKD 227

Qy     280 ELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLT 316
           || :|||:   :|   | |: |:||: :|:| |||||
Db     228 ELLNADEVFITSTTAEVIPVTSIDGQTIGSGAPGPLT 264


RESULT 6
ILVE_PSEAE
ID   ILVE_PSEAE     STANDARD;      PRT;   307 AA.
AC   O86428;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Branched-chain amino acid aminotransferase (EC 2.6.1.42) (BCAT).
GN   ILVE OR PA5013.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAK;
RA   Franklund C.V., Coyne M.J., Goldberg J.B.;
RT   "Clustering of the lipopolysaccharide core genes, waaF, waaC, waaG,
RT   and waaP, in Pseudomonas aeruginosa.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC       oxopentanoate + L-glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Valine and isoleucine biosynthesis.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U63816; AAC33172.1; -.
DR   EMBL; AE004913; AAG08398.1; -.
DR   HSSP; P00510; 1A3G.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     160    160       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   CONFLICT     85    104       RAAVRENNLESAYIRPMVFY -> APPCARTTWKAPISARW
FT                                 CST (IN REF. 1).
FT   CONFLICT    124    124       A -> S (IN REF. 1).
SQ   SEQUENCE   307 AA;  34084 MW;  5CD5CC6CFCDE423E CRC64;

  Query Match             17.3%;  Score 304.5;  DB 1;  Length 307;
  Best Local Similarity   30.0%;  Pred. No. 3.9e-17;
  Matches   93;  Conservative   47;  Mismatches  117;  Indels   53;  Gaps   10;

Qy      46 GVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAH---VWHG----------NIFRLGDH 92
           || | :|| :   :|   :        ||:|: :   |: |             ||||  |
Db       8 GVIWYDGELVQWRDATTHV---------LTHTLHYGMGVFEGVRAYDTPQGTAIFRLQAH 58

Qy      93 LDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTV---TRGYGKR-KGEKD 148
            ||| | |  : :   ||:||: | |:  |  : |  :::   |   : | | |  | |
Db      59 TDRLFDSAHIMNMQIPYSRDEINEATRAAVRENNLESAYIRPMVFYGSEGMGLRASGLK- 117

Qy     149 LSKLTHQVYIYAIPYLWAFPPAEQIF------GTTAIVPRHVR----RAGRNTVDPTIKN 198
                 |::    : |     |:         |::    ||     ||  |
```

```
Db    118 -------VHVIIAAWSWGAYMGEEALQQGIKVRTSSFTRHHVNISMTRAKSNGA------ 164

Qy    199 YQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNA-LPGITR 257
          : :   |  ||   ||  |::|| :  |||| | |: |:||| : :|   | | ||||
Db    165 --YINSMLALQEAISGGADEAMMLDPEGYVAEGSGENIFIIKDGVIYTPEVTACLNGITR 222

Qy    258 KTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTV 317
          |: || : | :   : :|  |:| |||     ||  ||||  :|| :| |  ||:|
Db    223 NTILTLAAEHGFKLVEKRITRDEVYIADEAFFTGTAAEVTPIREVDGRKIGAGRRGPVTE 282

Qy    318 AIRDRFWALM 327
            ::  :: |:
Db    283 KLQKAYFDLV 292


RESULT 7
ILVE_SALTY
ID   ILVE_SALTY     STANDARD;      PRT;   308 AA.
AC   P15168; Q9L6T0;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Branched-chain amino acid aminotransferase (EC 2.6.1.42) (Transaminase
DE   B) (BCAT).
GN   ILVE OR STM3903 OR STMD1.89 OR STY3654.
OS   Salmonella typhimurium, and
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602, 601;
RN   [1]
RP   SEQUENCE.
RC   SPECIES=S.typhimurium;
RX   MEDLINE=89352621; PubMed=2669973;
RA   Feild M.J., Nguyen D.C., Armstrong F.B.;
RT   "Amino acid sequence of Salmonella typhimurium branched-chain amino
RT   acid aminotransferase.";
RL   Biochemistry 28:5306-5310(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [3]
RP   SEQUENCE OF 1-5.
RC   SPECIES=S.typhimurium;
RA   Randall R.R., Wallis M.H., Young G.J., Armstrong F.B.;
RT   "N-terminal sequence of branched-chain amino acid aminotransferase.";
RL   Fed. Proc. 38:325-325(1979).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhi; STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
CC   -!- FUNCTION: Acts on leucine, isoleucine and valine.
CC   -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC       oxopentanoate + L-glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Valine and isoleucine biosynthesis.
CC   -!- SUBUNIT: HOMOHEXAMER.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF233324; AAF33481.1; -.
DR   EMBL; AE008882; AAL22753.1; -.
DR   EMBL; AL627279; CAD09414.1; -.
DR   PIR; A05077; A05077.
DR   PIR; A34082; A34082.
DR   HSSP; P00510; 1A3G.
DR   StyGene; SG10181; ilvE.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 2.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   INIT_MET      0      0
FT   BINDING     159    159       PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   308 AA;  33921 MW;  8F826434F9588AAD CRC64;

  Query Match             17.1%;  Score 302.5;  DB 1;  Length 308;
  Best Local Similarity   31.6%;  Pred. No. 5.6e-17;
  Matches   91;  Conservative   46;  Mismatches  136;  Indels   15;  Gaps    8;

Qy     49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
          |  || :  |:||: :     :   :     :  :    :||  :|: || | |   |
Db      9 WFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLRDSAKIYR 68

Qy    105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
          | ||| |  :  :  : |  :::   |  |      |        | | | | |:
Db     69 FPVSQSIDELMEACRDVIRKNNLTSAYIRPLVFVG-DVGMGVNPPPGYTTDVIIAAFPW- 126

Qy    165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
          |:  || :  |  |:|     ||  ||:    |   ||    |  :  ||:  | :  |
Db    127 GAYLGAEALDQGIDAMVSSW-NRAAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 183

Qy    221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
           || :  ::|| | |:  |||| | :|  : :|||||||  : :|| ::|||   : ::
Db    184 ALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQVLSR 243

Qy    279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWAL 326
          || |||:    ||  :||: |:||  || |  ||:|  |:  |: |
Db    244 ESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGL 291


RESULT 8
ILVE_ECOLI
ID   ILVE_ECOLI     STANDARD;      PRT;   308 AA.
AC   P00510; Q47299;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Branched-chain amino acid aminotransferase (EC 2.6.1.42) (Transaminase
DE   B) (BCAT).
GN   ILVE OR B3770 OR Z5281 OR ECS4704.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=85289113; PubMed=3897211;
RA   Kuramitsu S., Ogawa T., Ogawa H., Kagamiyama H.;
RT   "Branched-chain amino acid aminotransferase of Escherichia coli:
RT   nucleotide sequence of the ilvE gene and the deduced amino acid
RT   sequence.";
RL   J. Biochem. 97:993-999(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=87174741; PubMed=3550695;
RA   Lawther R.P., Wek R.C., Lopes J.M., Pereira R., Taillon B.E.,
RA   Hatfield G.W.;
RT   "The complete nucleotide sequence of the ilvGMEDA operon of
RT   Escherichia coli K-12.";
RL   Nucleic Acids Res. 15:2137-2155(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=92358234; PubMed=1379743;
RA   Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT   "Analysis of the Escherichia coli genome: DNA sequence of the region
RT   from 84.5 to 86.5 minutes.";
RL   Science 257:771-778(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [6]
RP   PRELIMINARY SEQUENCE OF 1-80 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=80101081; PubMed=392469;
RA   Lawther R.P., Nichols B.P., Zurawski G., Hatfield G.W.;
RT   "The nucleotide sequence preceding and including the beginning of the
RT   ilvE gene of the ilvGEDA operon of Escherichia coli K12.";
RL   Nucleic Acids Res. 7:2289-2301(1979).
RN   [7]
RP   SEQUENCE OF 1-33 FROM N.A.
RX   MEDLINE=92235862; PubMed=1569580;
RA   Pagel J.M., Winkelman J.W., Adams C.W., Hatfield G.W.;
RT   "DNA topology-mediated regulation of transcription initiation from
RT   the tandem promoters of the ilvGMEDA operon of Escherichia coli.";
RL   J. Mol. Biol. 224:919-935(1992).
RN   [8]
RP   PYRIDOXAL-PHOSPHATE BINDING SITE.
RX   MEDLINE=89174510; PubMed=3069843;
RA   Inoue K., Kuramitsu S., Aki K., Watanabe Y., Takagi T., Nishigai M.,
RA   Ikaiu A., Kagamiyama H.;
RT   "Branched-chain amino acid aminotransferase of Escherichia coli:
RT   overproduction and properties.";
RL   J. Biochem. 104:777-784(1988).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=97306043; PubMed=9163511;
RA   Okada K., Hirotsu K., Sato M., Hyashi H., Kagamiyama H.;
RT   "Three-dimensional structure of Escherichia coli branched-chain amino
RT   acid aminotransferase at 2.5-A resolution.";
RL   J. Biochem. 121:637-641(1997).
CC   -!- FUNCTION: Acts on leucine, isoleucine and valine.
CC   -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC       oxopentanoate + L-glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Valine and isoleucine biosynthesis.
CC   -!- SUBUNIT: HOMOHEXAMER.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10313; AAB59052.1; -.
DR   EMBL; X04890; CAA28575.1; -.
DR   EMBL; M87049; AAA67573.1; -.
DR   EMBL; AE000453; AAC77490.1; -.
DR   EMBL; AE005608; AAG58965.1; -.
DR   EMBL; AP002566; BAB38127.1; -.
DR   EMBL; X02413; CAA26262.1; -.
DR   EMBL; M32253; AAA24022.1; -.
DR   EMBL; V00290; CAA23559.1; -.
DR   PIR; A00602; XNECV.
DR   PIR; S30668; S30668.
DR   PDB; 1A3G; 27-MAY-98.
DR   SWISS-2DPAGE; P00510; COLI.
DR   ECO2DBASE; F032.5; 6TH EDITION.
DR   EcoGene; EG10497; ilvE.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; 3D-structure; Complete proteome.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       THR MODIFIED TO GLY.
FT   BINDING     159    159       PYRIDOXAL PHOSPHATE.
FT   CONFLICT    150    150       A -> R (IN REF. 3).
SQ   SEQUENCE   308 AA;  33962 MW;  28404978D2B79659 CRC64;

  Query Match             16.9%;  Score 298.5;  DB 1;  Length 308;
  Best Local Similarity   30.9%;  Pred. No. 1.2e-16;
  Matches   89;  Conservative   49;  Mismatches  135;  Indels   15;  Gaps    8;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
           |  ||  :  |:||: :      :    :     :  :    :||  :|: || | |   |
Db       9 WFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYDSHKGPVVFRHREHMQRLHDSAKIYR 68

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           | ||| |  :  :  : |  :::   :  |     |    :  :  | | | |:
Db      69 FPVSQSIDELMEACRDVIRKNNLTSAYIRPLIFVG-DVGMGVNPPAGYSTDVIIAAFPW- 126

Qy     165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
           |:  || :  |  |:|     ||  ||:    |   ||    |  :  ||:  | :  |
Db     127 GAYLGAEALEQGIDAMVSSW-NRAAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 183

Qy     221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
           || :  ::|| | |:  |||| | :|  : :|||||||  : :|| ::|||   : ::
Db     184 ALDVNGYISEGAGENLFEVKDGVLFTPPFTSSALPGITRDAIIKLAKELGIEVREQVLSR 243

Qy     279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWAL 326
           || |||:    ||  :||: |:||  || |  ||:|  |:  |: |
```

```
Db    244 ESLYLADEVFMSGTAAEITPVRSVDGIQVGEGRCGPVTKRIQQAFFGL 291


RESULT 9
DAAA_LISMO
ID   DAAA_LISMO     STANDARD;      PRT;   289 AA.
AC   O85046;
DT   30-MAY-2000 (Rel. 39, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
DE   aminotransferase) (D-amino acid aminotransferase) (D-amino acid
DE   transaminase) (DAAT).
GN   DAT OR DAAA OR LMO1619.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=10403S;
RX   MEDLINE=98339852; PubMed=9673233;
RA   Thompson R.J., Bouwer H.G.A., Portnoy D.A., Frankel F.R.;
RT   "Pathogenicity and immunogenicity of a Listeria monocytogenes strain
RT   that requires D-alanine for growth.";
RL   Infect. Immun. 66:3552-3561(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-e / Serovar 1/2a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
CC   -!- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,
CC       GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-
CC       glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF038439; AAC31363.1; -.
DR   EMBL; AL591980; CAC99697.1; -.
DR   HSSP; P19938; 1DAA.
DR   ListiList; LMO01619; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; FALSE_NEG.
KW   Transferase; Aminotransferase; Pyridoxal phosphate; Complete proteome.
FT   BINDING     147    147       PYRIDOXAL PHOSPHATE (POTENTIAL).
FT   CONFLICT    286    286       V -> E (IN REF. 1).
SQ   SEQUENCE   289 AA;  32402 MW;  54A144692E7FEE35 CRC64;
```

```
  Query Match           16.4%;  Score 289;  DB 1;  Length 289;
  Best Local Similarity  28.9%;  Pred. No. 6.2e-16;
  Matches   83;  Conservative   51;  Mismatches  123;  Indels   30;  Gaps    6;

Qy     50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109
          :    : |:| : | | |:   |  | |  :::|  |   :|:|||   |:|: |   |
Db      5 VNNHLVEREDATVDIEDRGYQFGDGVYEVVRLYNGKFFTYNEHIDRLYASAAKIDLVIPY 64

Qy    110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPP 169
          ||:|| |: :| |: : :    | | ||||    :              : ||    ||
Db     65 SKEELRELLEKLVAENNINTGNVYLQVTRGVQNPRN------------HVIPD--DFPL 109

Qy    170 AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGART 218
             :      |||: |:        | :: :|           |:: | : :|  : |
Db    110 EGVLTAAAREVPRNERQFVEGGTAITEEDVRWLRCDIKSLNLLGNILAKN-KAHQQNALE 168

Qy    219 AILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALPGITRKTVFELADQMGIEATLRDV 276
          |||   :  | |    || |:||| | : : :   | ||||: : ::| : ||     |
Db    169 AILHRGEQ-VTECSASNVSIIKDGVLWTHAADNLILNGITRQVIIDVAKKNGIPVKEADF 227

Qy    277 TSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
          |  :| :|||:   :|   :||| :||  | :|: ||:|  :   |
Db    228 TLTDLREADEVFISSTTIEITPITHIDGVQVADGKRGPITAQLHQYF 274


RESULT 10
DAAA_STAHA
ID   DAAA_STAHA     STANDARD;      PRT;   282 AA.
AC   P54694;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
DE   aminotransferase) (D-amino acid aminotransferase) (D-amino acid
DE   transaminase) (DAAT).
GN   DAT.
OS   Staphylococcus haemolyticus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1283;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y176;
RX   MEDLINE=95113770; PubMed=7814322;
RA   Pucci M.J., Thanassi J.A., Ho H.-T., Falk P.J., Dougherty T.J.;
RT   "Staphylococcus haemolyticus contains two D-glutamic acid
RT   biosynthetic activities, a glutamate racemase and a D-amino acid
RT   transaminase.";
RL   J. Bacteriol. 177:336-342(1995).
CC   -!- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,
CC       GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-
CC       glutamate.
CC   -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12238; AAA20396.1; -.
DR   HSSP; P19938; 1DAA.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Pyridoxal phosphate.
```

```
FT   BINDING     146    146       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   282 AA;  31769 MW;  43DEC87A570B4F13 CRC64;

  Query Match             16.0%;  Score 283;  DB 1;  Length 282;
  Best Local Similarity   25.3%;  Pred. No. 1.8e-15;
  Matches   73;  Conservative   59;  Mismatches  124;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :| ||::   |||:|  | |:   |  |     : | :| : :| :|  :  ||:::|| |
Db       5 FINGEFIDQNEAKVSYEDRGYVFGDGIYEYIRAYDGKLFTVTEHFERFIRSASEIQLDLG 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           |: :|| :: :: : :: ::   : :  |||  |       ::   :  :|  |   :
Db      65 YTVEELIDVVRELLKVNNIQNGGIYIQATRGVAPRNHSFPTPEVKPVIMAFAKSYDRPYD 124

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGAR 217
             |                 | |    |::: :|          |:: |  :  |
Db     125 DLEN--------------GINAA--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAGE 167

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
              :      | ||   ||  :||| :   |  |  | |||||  :  :::   |
Db     168 A--IQHRGETVTEGASSNVYAIKDGAIYTHPVNNYILNGITRKVIKWISEDEDIPFKEET 225

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |   | :|||::  :|:  |||:  :||| ||:|: ||:|  ::: |
Db     226 FTVEFLKNADEVIVSSTSAEVTPVVKIDGEQVGDGKVGPVTRQLQEGF 273


RESULT 11
DAAA_BACSP
ID   DAAA_BACSP     STANDARD;      PRT;   282 AA.
AC   P19938;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
DE   aminotransferase) (D-amino acid aminotransferase) (D-amino acid
DE   transaminase) (DAAT).
GN   DAT.
OS   Bacillus sp. (strain YM-1).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1409;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89123328; PubMed=2644261;
RA   Tanizawa K., Asano S., Masu Y., Kuramitsu S., Kagamiyama H.,
RA   Tanaka H., Soda K.;
RT   "The primary structure of thermostable D-amino acid aminotransferase
RT   from a thermophilic Bacillus species and its correlation with L-amino
RT   acid aminotransferases.";
RL   J. Biol. Chem. 264:2450-2454(1989).
RN   [2]
RP   SEQUENCE OF 1-20; 142-156 AND 280-282.
RX   MEDLINE=89123327; PubMed=2914916;
RA   Tanizawa K., Masu Y., Asano S., Tanaka H., Soda K.;
RT   "Thermostable D-amino acid aminotransferase from a thermophilic
RT   Bacillus species. Purification, characterization, and active site
RT   sequence determination.";
RL   J. Biol. Chem. 264:2445-2449(1989).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.94 ANGSTROMS).
RX   MEDLINE=95352651; PubMed=7626635;
RA   Sugio S., Petsko G.A., Manning J.M., Soda K., Ringe D.;
RT   "Crystal structure of a D-amino acid aminotransferase: how the
RT   protein controls stereoselectivity.";
RL   Biochemistry 34:9661-9669(1995).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE=98206914; PubMed=9538014;
RA   Peisach D., Chipman D.M., van Ophem P.W., Manning J.M., Petsko G.A.,
RA   Ringe D.;
RT   "Crystallographic study of steps along the reaction pathway of
RT   D-amino acid aminotransferase.";
RL   Biochemistry 37:4958-4967(1998).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=98420361; PubMed=9749913;
RA   Sugio S., Kashima A., Kishimoto K., Peisach D., Petsko G.A., Ringe D.,
RA   Yoshimura T., Esaki N.;
RT   "Crystal structures of L201A mutant of D-amino acid aminotransferase
RT   at 2.0-A resolution: implication of the structural role of Leu201 in
RT   transamination.";
RL   Protein Eng. 11:613-619(1998).
CC   -!- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,
CC       GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE.
CC   -!- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-
CC       glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04460; AAA22252.1; -.
DR   PIR; A31422; A31422.
DR   PIR; A31421; A31421.
DR   PDB; 1DAA; 15-SEP-95.
DR   PDB; 2DAA; 18-MAR-98.
DR   PDB; 3DAA; 29-APR-98.
DR   PDB; 4DAA; 29-APR-98.
DR   PDB; 5DAA; 16-MAR-99.
DR   PDB; 1A0G; 03-JUN-98.
DR   PDB; 2DAB; 03-JUN-98.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Pyridoxal phosphate; 3D-structure.
FT   INIT_MET      0      0
FT   BINDING     145    145       PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   282 AA;  32265 MW;  E40F28C79417B1BE CRC64;

  Query Match             15.8%;  Score 278.5;  DB 1;  Length 282;
  Best Local Similarity   27.6%;  Pred. No. 4.1e-15;
  Matches   83;  Conservative   50;  Mismatches  131;  Indels   37;  Gaps    9;

Qy      45 GGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLR 104
           |  .|   : :  || ||   | |:   |  | |  |::| :| : :|:|||   | |:|
Db       1 GYTLW-NDQIVKDEEVKIDKEDRGYQFGDGVYEVVKVYNGEMFTVNEHIDRLYASAEKIR 59

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           :   |:||:  ::  : |  ::|    :   ||||  |          ||
Db      60 ITIPYTKDKFHQLLHELVEKNELNTGHIYFQVTRGTSPR---------AHQ--------- 101

Qy     165 WAFPP---AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW--GDLTA--------ASFEA 211
             ||       | | |   || :    :       ::: :|   |: :          |  ||
Db     102 --FPENTVKPVIIGYTKENPRPLENLEKGVKATFVEDIRWLRCDIKSLNLLGAVLAKQEA 159

Qy     212 KDRGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGI 269
           ::|   || |  :| | ||   ||  :||| |   |: |  | ||||  |   |::: :
Db     160 HEKGCYEAI-LHRNNTVTEGSSSNVFGIKDGILYTHPANNMILKGITRDVVIACANEINM 218

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
                 |: |    |||   :|   :||:  :||: : :|: |  |  :: :|   : :
Db     219 PVKEIPFTTHEALKMDELFVTSTTSEITPVIEIDGKLIRDGKVGEWTRKLQKQFETKIPK 278

Qy     330 P 330
           |
```

```
Db    279 P 279


RESULT 12
DAAA_BACSU
ID   DAAA_BACSU     STANDARD;      PRT;   282 AA.
AC   O07597;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
DE   aminotransferase) (D-amino acid aminotransferase) (D-amino acid
DE   transaminase) (DAAT).
GN   DAT.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,
CC       GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-
CC       glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y14082; CAA74512.1; -.
DR   EMBL; Y14080; CAA74451.1; -.
DR   EMBL; Z99109; CAB12806.1; -.
DR   HSSP; P19938; 2DAB.
DR   SubtiList; BG13045; dat.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Pyridoxal phosphate; Complete proteome.
FT   BINDING     144    144       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   282 AA;  31182 MW;  49D5B75E5DF7DEC1 CRC64;

  Query Match             15.8%;  Score 278.5;  DB 1;  Length 282;
  Best Local Similarity   29.1%;  Pred. No. 4.1e-15;
  Matches   82;  Conservative   48;  Mismatches  129;  Indels   23;  Gaps    6;

Qy      50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109
           : |  :   || | : | |:   |  | |  |: | :| | :| :|    |::: :   :
Db       5 VNGRLIGRSEASIDLEDRGYQFGDGIYEVIRVYKGVLFGLREHAERFFRSAAEIGISLPF 64

Qy     110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAF-- 167
           | ::|    :| |  : : |  | :  |||   ||          ||      |   |:
Db      65 SIEDLEWDLQKLVQENAVSEGAVYIQTTRGVAPRK---------HQYEAGLEPQTTAYTF 115

Qy     168 ----PPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
               |  || :|  ||    :|       |   | :  :|     | : ||  ||||
Db     116 TVKKPEQEQAYGVAAITDEDLRWL---RCDIKSLNLLYNVMTKQ--RAYEAGAFEAILL- 169

Qy     224 SDNCVAEGPGFNVCIVKDGKLAS--PSRNALPGITRKTVFELADQMGIEATLRDVTSREL 281
             |  | ||   ||  | :| : :   :|  | ||||  :  | :: ||:     |:  ||
Db     170 RDGVVTEGTSSNVYAVINGTVRTHPANRLILNGITRMNILGLIEKNGIKLDETPVSEEEL 229

Qy     282 YDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
             |:|:   :|   : |: :|||:::|:|:|||:|  ::  |
Db     230 KQAEEIFISSTTAEIIPVVTLDGQSIGSGKPGPVTKQLQAAF 271


RESULT 13
DAAA_LISIN
ID   DAAA_LISIN     STANDARD;      PRT;   289 AA.
AC   Q92B90;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
DE   aminotransferase) (D-amino acid aminotransferase) (D-amino acid
DE   transaminase) (DAAT).
GN   DAT OR DAAA OR LIN1660.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / Serovar 6a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
CC   -!- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,
CC       GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE (By
```

```
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-
CC       glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL596169; CAC96891.1; -.
DR   ListiList; LIN01660; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; FALSE_NEG.
KW   Transferase; Aminotransferase; Pyridoxal phosphate; Complete proteome.
FT   BINDING     147    147       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   289 AA;  32323 MW;  30B05F95138F4173 CRC64;

  Query Match             15.0%;  Score 264;  DB 1;  Length 289;
  Best Local Similarity   28.1%;  Pred. No. 6.1e-14;
  Matches   78;  Conservative   51;  Mismatches  119;  Indels   30;  Gaps    6;

Qy      59 EAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEIT 118
           :| : : | |:   |  | |  :::|  |   :|:|||   |:|: |   |||: |  :
Db      14 DATVDVEDRGYQFGDGVYEVVRLYNGKFFTYNEHIDRLYASAAKIDLVIPYSKETLRALL 73

Qy     119 KKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTA 178
           | |: : :    | | ||||    :               : :|    ||    :
Db      74 DKLVAENNINTGNVYLQVTRGVQNPRN-------------HVLPD--DFPLEGVLTAAAR 118

Qy     179 IVPRHVRRAGRNTVDPTIKNYQW-----------GDLTAASFEAKDRGARTAILLDSDNC 227
           |||: |:        | :: :|           |:: | : :|  : |  |||   :
Db     119 EVPRNERQFIEGGSAITEEDVRWLRCDIKSLNLLGNILAKN-KAHQQNALEAILHRGEQ- 176

Qy     228 VAEGPGFNVCIVKDGKLASPSRN--ALPGITRKTVFELADQMGIEATLRDVTSRELYDAD 285
           . | |    || |:||| | : : :   | ||||: : ::| : ||     | |  :| :||
Db     177 VTECSASNVSIIKDGVLWTHAADNLILNGITRQVIIDVAKKNGIPVKEADFTLTDLREAD 236

Qy     286 ELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |:   :|   :|||  :||  | :|: ||:|  : : |
Db     237 EVFISSTTIEITPITHIDGVQVADGKRGPITAQLHNYF 274


RESULT 14
DAAA_BACSH
ID   DAAA_BACSH     STANDARD;      PRT;   283 AA.
AC   P54693;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
DE   aminotransferase) (D-amino acid aminotransferase) (D-amino acid
DE   transaminase) (DAAT).
GN   DAT.
OS   Bacillus sphaericus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1421;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10208;
RA   Fotheringham I.G., Taylor P.P., Bledig S.A.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 28;
RA   Hwang T.S., Wang N.C., Wu S.P., Yang C.L., Tsai H.;
RT   "Cloning and overexpression of D-amino acid aminotransferase from
RT   Bacillus sphaericus in Escherichia coli.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,
CC       GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-
CC       glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U26732; AAA68028.1; -.
DR   EMBL; AF081278; AAC33964.1; -.
DR   HSSP; P19938; 1DAA.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Pyridoxal phosphate.
FT   BINDING     146    146       PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ   SEQUENCE   283 AA;  31477 MW;  6FADA97D5CA01A8D CRC64;

  Query Match             14.2%;  Score 250;  DB 1;  Length 283;
  Best Local Similarity   26.9%;  Pred. No. 7.7e-13;
  Matches   79;  Conservative   49;  Mismatches  110;  Indels   56;  Gaps    9;

Qy      58 EEAKISIF--DTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELA 115
           ||  |:|   | |:   |  | |  |::|::|   :|:||    | |:||   |:|| |
Db      12 EEGSITISPEDRGYQFGDGIYEVIKVYNGHMFTAQEHIDRFYASAEKIRLVIPYTKDVLH 71

Qy     116 EITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFG 175
           ::    :  : |    |   :|||   |                      || |
Db      72 KLLHDLIEKNNLNTGHVYFQITRGTTSRN-------------------HIFPDA----S 107

Qy     176 TTAIVPRHVRRAGRNTVDPTIKNYQWG-------------DLTA--------ASFEAKD 213
             |::  :|:     | : :|:|:: |              |: :        |  || :
Db     108 VPAVLTGNVK-----TGERSIENFEKGVKATLVEDVRWLRCDIKSLNLLGAVLAKQEASE 162

Qy     214 RGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEA 271
           :|   |||   |  : |    ||  :|||||   |: |  | ||||: : : | :: :
Db     163 KGCYEAILHRGD-IITECSSANVYGIKDGKLYTHPANNYILNGITRQVILKCAAEINLPV 221

Qy     272 TLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWA 325
           :|  :|   ||::  : :  |||:  :||: :| | ||  |  ::  | |
Db     222 IEEPMTKGDLLTMDEIIVSSVSSEVTPVIDVDGQQIGAGVPGEWTRKLQKAFEA 275


RESULT 15
ILVE_RICCN
ID   ILVE_RICCN     STANDARD;      PRT;   290 AA.
AC   Q92I26;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable branched-chain amino acid aminotransferase (EC 2.6.1.42)
DE   (BCAT).
GN   ILVE OR RC0594.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
CC   -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC       oxopentanoate + L-glutamate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Valine and isoleucine biosynthesis.
CC   -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008620; AAL03132.1; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW   Pyridoxal phosphate; Complete proteome.
FT   BINDING     155    155       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   290 AA;  33243 MW;  86112A254A505177 CRC64;

  Query Match             12.0%;  Score 212;  DB 1;  Length 290;
  Best Local Similarity   24.7%;  Pred. No. 8.6e-10;
  Matches   67;  Conservative   52;  Mismatches  124;  Indels   28;  Gaps    5;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           || |: :| : |:| :      :|   :     ::| :|:| :|  ||:  |  | |
Db      14 WINGDLVPYQFARIHVLTHSLHYSGSVFEGERAYNGKVFKLKEHTARLIKSAEALGLKVP 73

Qy     109 YSKDELAEITKKCVSMSQLRESFV---------NLTVTRGYGKRKGEKDLSKLTHQVYIY 159
           |: ||: :  :  :  : ::::::         :| :|  |            |:  : |
Db      74 YNVDEIIKAHECVIKQNNIKDAYIRPLIWCGDESLNITNQY----------LSTNLLIA 122

Qy     160 AIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTA 219
            || :    |     |    | |  :    :|   :    |:     :  |||  |   |
Db     123 GIPSM----PRSFEKGINLHVSRWRKAMPDSTPVQSKSAAQYNMAITSKKEAKALGYEDA 178

Qy     220 ILLDSDNCVAEGPGFNVCIVKDGKLASP-SRNALPGITRKTVFELADQMGIEATLRDVTS 278
           :||| :  :||    |:  |||  | :| :   | ||||:|: |:|  :|:|     :
Db     179 LLLDYEGYIAECTTTNIFFVKDKILYTPIADRFLNGITRQTIIEIAKDLGLEVKEERLKL 238

Qy     279 RELYDADELMAVTTAGGVTPINSLDGEAVGN 309
            :: |        ||  |  |:|:|   :||
Db     239 EQIEDFTGCFVTGTAIEVQNIDSID---LGN 266


Search completed: April 17, 2003, 17:37:55
Job time : 27 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:35:22 ; Search time 88 Seconds
(without alignments)
793.750 Million cell updates/sec

Title: US-10-067-291-1
Perfect score: 1764
Sequence: 1 MTALSDLGTSNLVAVEPGAI.....RDRFWALMDEPGPLIETIEY 339

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	596.5	33.8	320	16	Q98K82	Q98k82 rhizobium l
2	577	32.7	291	16	Q981W8	Q981w8 rhizobium l
3	414.5	23.5	295	17	Q8TUX5	Q8tux5 methanopyru
4	394	22.3	294	17	Q8TI08	Q8ti08 methanosarc
5	352.5	20.0	291	16	Q8RB89	Q8rb89 thermoanaer
6	340.5	19.3	555	10	Q8W0Z7	Q8w0z7 arabidopsis
7	326.5	18.5	559	10	Q9ASR4	Q9asr4 arabidopsis
8	318.5	18.1	490	10	Q9MAA0	Q9maa0 arabidopsis
9	315	17.9	303	17	Q8ZTE2	Q8zte2 pyrobaculum
10	301	17.1	308	16	Q8ZAB2	Q8zab2 yersinia pe
11	298	16.9	308	17	Q9HNF8	Q9hnf8 halobacteri
12	291	16.5	282	2	Q9KWZ6	Q9kwz6 staphylococ
13	290	16.4	282	16	Q99TB4	Q99tb4 staphylococ
14	289.5	16.4	290	16	Q8XRW7	Q8xrw7 ralstonia s
15	278	15.8	287	16	Q92Q85	Q92q85 rhizobium m
16	277.5	15.7	319	16	Q9KVV9	Q9kvv9 vibrio chol
17	273.5	15.5	296	16	Q9A4A8	Q9a4a8 caulobacter
18	273	15.5	293	16	Q8YD14	Q8yd14 brucella me
19	265.5	15.1	309	16	Q8Y1X0	Q8y1x0 ralstonia s
20	263	14.9	289	2	Q937U0	Q937u0 brucella su
21	260	14.7	283	16	Q9K939	Q9k939 bacillus ha
22	259	14.7	319	16	Q8UCH6	Q8uch6 agrobacteri
23	256	14.5	287	16	Q98MX3	Q98mx3 rhizobium l
24	253	14.3	304	16	Q9PIM6	Q9pim6 campylobact
25	247.5	14.0	297	16	Q98HX8	Q98hx8 rhizobium l
26	246	13.9	290	16	Q8UJK4	Q8ujk4 agrobacteri
27	244.5	13.9	285	17	Q9HS64	Q9hs64 halobacteri
28	240	13.6	287	16	Q9A7H9	Q9a7h9 caulobacter
29	238.5	13.5	283	16	Q930Z7	Q930z7 rhizobium m
30	223.5	12.7	296	10	Q9FJM9	Q9fjm9 arabidopsis
31	217	12.3	290	2	Q9AKK0	Q9akk0 rickettsia
32	212.5	12.0	295	10	Q9FTU3	Q9ftu3 oryza sativ
33	208	11.8	290	2	Q9AKQ4	Q9akq4 rickettsia
34	205.5	11.6	280	16	Q97KX4	Q97kx4 clostridium
35	204.5	11.6	271	16	Q9HZN6	Q9hzn6 pseudomonas
36	199	11.3	303	2	Q93H01	Q93h01 streptomyce
37	197.5	11.2	288	2	Q9AKE5	Q9ake5 rickettsia
38	178.5	10.1	359	2	Q9F329	Q9f329 staphylococ
39	176.5	10.0	273	16	Q9L1D6	Q9l1d6 streptomyce
40	172.5	9.8	292	16	Q985K4	Q985k4 rhizobium l
41	165.5	9.4	217	2	Q9AM72	Q9am72 rhodospiril
42	164.5	9.3	410	10	Q9M2Z0	Q9m2z0 arabidopsis
43	161.5	9.2	418	10	Q9SNY9	Q9sny9 solanum tub
44	161	9.1	276	2	Q9F7N1	Q9f7n1 uncultured
45	160.5	9.1	295	16	Q9KGH0	Q9kgh0 bacillus ha

ALIGNMENTS

RESULT 1
Q98K82
ID Q98K82 PRELIMINARY; PRT; 320 AA.
AC Q98K82;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Branched-chain amino acid transferase.
GN MLR1594.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002997; BAB48932.1; -.
DR InterPro; IPR001544; Aminotran_4.
DR Pfam; PF01063; aminotran_4; 1.
DR ProDom; PD001961; Aminotran_4; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 320 AA; 34742 MW; 837EAFACACA9F614 CRC64;

Query Match 33.8%; Score 596.5; DB 16; Length 320;
Best Local Similarity 40.7%; Pred. No. 3.6e-40;
Matches 121; Conservative 57; Mismatches 116; Indels 3; Gaps 3;

```
Qy      34 DYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHL 93
           |  :|  | :  |:|:::|:|||  :||:|: | || ||| ||   |||:|  |||  ||
Db      18 DRHVDPHS-YPDGIAFLDGQYLPMSQAKVSVLDWGFLHSDATYDTVHVWNGRFFRLDLHL 76
```

```
Qy      94 DRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLT 153
           ||   |  |||:   : :| :|||   ||::|  | ::| :  |||       :|  :
Db      77 DRFFGGLEKLRMTIPFDRDGVAEILHNCVALSGHRAAYVEMLCTRG-ASPTFSRDPRQAI 135

Qy     154 HQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKD 213
           ::   :|:|:  :   |||:     :      |    :|||:|||| | ||    ::| |
Db     136 NRFMAFAVPF-GSVANAEQLQRGLRVAISDKVRIPPASVDPSIKNYHWLDLVRGLYDAYD 194

Qy     214 RGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGIEATL 273
           || ||::|| :  ||||||||  ||||||::|:   ||||||:|||:|  : |: |
Db     195 SGAETALILDFNGNVAEGPGFNVFCVKDGKLSTPAIGVLPGITRRTVFDLCAEEGLAAAA 254

Qy     274 RDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEP 330
           ||:   |  |||:   :||||: |:  :|| |: :|: ||:|  :   :|   |:|
Db     255 ADVSVAALKAADEVFITSTAGGIMPVTEIDGAAIADGKVGPVTSRLMALYWQKHDDP 311


RESULT 2
Q981W8
ID   Q981W8      PRELIMINARY;      PRT;   291 AA.
AC   Q981W8;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Branched-chain amino acid aminotransferase.
GN   MLL9205.
OS   Rhizobium loti (Mesorhizobium loti).
OG   Plasmid pMLa.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003015; BAB54591.1; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
KW   Transferase; Aminotransferase; Plasmid; Complete proteome.
SQ   SEQUENCE   291 AA;  32309 MW;  E7981CD1EABB3EF1 CRC64;

  Query Match           32.7%;  Score 577;  DB 16;  Length 291;
  Best Local Similarity  39.4%;  Pred. No. 1.2e-38;
  Matches  112;  Conservative   60;  Mismatches  110;  Indels    2;  Gaps    2;

Qy      47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           :|:::|:|||  |||:|: | || ||| ||   ||| |  |||  |:||   |  |||:
Db       1 MAFMDGQYLPMSEAKVSVLDWGFLHSDATYDTVHVWDGRFFRLNLHVDRFFRGMEKLRMK 60

Qy     107 AGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWA 166
             |:: |: :|   ||::|  : ::| :  ||| |     :|  :  ::   :|:|:  :
Db      61 LPYNRSEIEKILSTCVALSGHKSAYVEMICTRG-GSPTFSRDPRQSENRFIAFAVPF-GS 118

Qy     167 FPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDN 226
             ||:     :   :  |    ::|||||| | ||    |:| | || ||:::| ::
Db     119 VANKEQLERGLHVAISNTVRIPPKSIDPTIKNYHWLDLVKGLFDAYDYGAETALIVDIND 178

Qy     227 CVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADE 286
           :|||||||  ||||:| :|:   | ||||:|||:| |::|:  :  |:   ||  |||
Db     179 NIAEGPGFNVFTVKDGRLKTPAYGVLAGITRQTVFDLCDELGLSVSAGDIDRNELKGADE 238

Qy     287 LMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEP 330
           :   :||||: |:: :|   ||:|: | ||  : | :|    :|
Db     239 VFITSTAGGIMPVSKIDETVVGDGKVGALTRQLADLYWEKHADP 282


RESULT 3
Q8TUX5
ID   Q8TUX5      PRELIMINARY;      PRT;   295 AA.
AC   Q8TUX5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Branched-chain amino acid aminotransferase.
GN   ILVE OR MK1627.
OS   Methanopyrus kandleri.
OC   Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC   Methanopyrus.
OX   NCBI_TaxID=2320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AV19 / DSM 6324 / JCM 9639;
RX   MEDLINE=21927647; PubMed=11930014;
RA   Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA   Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA   Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA   Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT   "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT   and monophyly of archaeal methanogens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR   EMBL; AE010451; AAM02840.1; -.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   295 AA;  32955 MW;  5DB962E995D549BB CRC64;

  Query Match           23.5%;  Score 414.5;  DB 17;  Length 295;
  Best Local Similarity  33.0%;  Pred. No. 1.8e-25;
  Matches   97;  Conservative   60;  Mismatches  122;  Indels   15;  Gaps    4;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: || :| |||||::| || : |  :      : | ||:| :|:||| | |  : |:
Db       9 YLNGELVPREEAKISVYDHGFLYGDGVFEGIRAYDGRIFKLDEHVDRLYDSAKAIMLEIP 68

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           :|::: |   : |  ::||:::: : |:|| |    : :     : | |
Db      69 MTKEKMKEAIIETVRANELRDAYIRVVVSRGEGDLGLDPEKCPEPNVVII--------AE 120

Qy     169 PAEQIFGT-----TAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
           | | ::|        ::   |||   : :|| ||:  : :   |  :|   ||  ||:||
Db     121 PMEPLYGDLYEKGIEVITASVRRIPPDALDPKIKSCNYLNNILAKIQANLAGADEAIMLD 180

Qy     224 SDNCVAEGPGFNVCIVKDGKLASPSRNA-LPGITRKTVFELADQMGIEATLRDVTSRELY 282
           :  | || | || :|:|| : :|  :  | |||| || |: :::||    : :|  |||
Db     181 HEGYVCEGTGDNVFVVEDGTVYTPPEDTILRGITRATVMEICEELGIPVEEKRITLGELY 240

Qy     283 DADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPG-PLIE 335
           |||:    ||  | |:  :||  :|   |||:|  | : |  |  : | |: |
Db     241 AADEVFLTGTAAEVAPVRKVDGRKIGEECPGPITRRIMEAFRELTKKEGTPVYE 294


RESULT 4
Q8TI08
ID   Q8TI08      PRELIMINARY;      PRT;   294 AA.
AC   Q8TI08;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Branched chain amino acid aminotransferase.
GN   ILVE OR MA4349.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
```

```
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
DR   EMBL; AE011153; AAM07692.1; -.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   294 AA;  32240 MW;  D399F73A38973A62 CRC64;

  Query Match             22.3%;  Score 394;  DB 17;  Length 294;
  Best Local Similarity   31.2%;  Pred. No. 8e-24;
  Matches   91;  Conservative   65;  Mismatches  122;  Indels   14;  Gaps    5;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: |:::| :|||:|::| || : |  :     ::| :|:| :|:||| | |  : :|
Db       9 YLNGKFVPKDEAKVSVYDHGFLYGDGVFEGIRAYNGRVFKLKEHVDRLYDSAKAIAMDIP 68

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYG-----KRKGEKDLSKLTHQVYIYAIPY 163
           :::|: ||  : :  : |:::::   |:|| |       ||  |         | : |  :
Db      69 VTREEMTEIILEALRKNNLKDAYIRPIVSRGIGDLGLDPRKCGKP------SVIVIAQGW 122

Qy     164 LWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
           :    :: | |  |   |||   : : | ||:  : :   |  || ::|   || ||
Db     123 GAMYGDLYEV-GLTG-VSVCVRRNAPDALSPNIKSLNYLNNILAKIEANEKGGDEAIFLD 180

Qy     224 SDNCVAEGPGFNVCIVKDGKLASP-SRNALPGITRKTVFELADQMGIEATLRDVTSRELY 282
           :  | || | |: :||| :: :| : : | |||| |  || |:|| :    ::   :||
Db     181 HNGFVCEGSGDNIFVVKDDRVLTPFTISNLKGITRATAIELLDEMGYKVIESNLGLFDLY 240

Qy     283 DADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLI 334
             |||:    ||    |:  |||  :| |:|||||: : : |  :    |  |
Db     241 TADEIFVTGTAAESAPVTRLDGRVIGTGKPGPLTMKMVEAFAKITQSTGTQI 292


RESULT 5
Q8RB89
ID   Q8RB89      PRELIMINARY;      PRT;   291 AA.
AC   Q8RB89;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate
DE   lyase.
GN   ILVE OR TTE0933.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013059; AAM24189.1; -.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   291 AA;  32346 MW;  105072D27AB06FFE CRC64;

  Query Match             20.0%;  Score 352.5;  DB 16;  Length 291;
  Best Local Similarity   31.2%;  Pred. No. 1.8e-20;
  Matches   91;  Conservative   55;  Mismatches  117;  Indels   29;  Gaps    5;

Qy      46 GVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRL 105
           |: :: ||:: :|:| :|:|| |: : |  :     |: | ||:| :|| ||   |  | |
Db       2 GLVYVNGEFVDSEKACVSVFDHGYLYGDGIFEGIRVYDGVIFKLDEHLKRLYSMAKVLLL 61

Qy     106 DAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLW 165
           |   | :|: |   : | :: ||:::: | |:||    ||:  |             ||
Db      62 DIPLSMEEMKEKVVETVRINNLRDAYIRLVVSRG----KGDLGLD----------PYKC 106

Qy     166 AFPPAEQIFGTTAIVPRHVRRAG---------RNTV---DPTIKNYQWGDLTAASFEAKD 213
           |    |   |: |  :   |          ||::   || ||:  : :   |  ||
Db     107 PKPTVVIIADKIALYPESMYETGLKIITSSFRRNSIQSLDPQIKSLNYLNNILAKIEAVK 166

Qy     214 RGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEA 271
           |   |::|: :  |||  | |: |: :| | ||  :  || |||| || ::| :: |
Db     167 AGYPEALILNQEGYVAECTGDNIFIISNGVLYSPPSAVGALGGITRATVIDIARELNIPF 226

Qy     272 TLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           : :   :| |||      ||    |:  :|   :|:|:|||:|  | : |
Db     227 EEKYFSLFNVYTADECFLTGTAAEAIPVVEVDHRVIGDGKPGPITKKIIEEF 278


RESULT 6
Q8W0Z7
ID   Q8W0Z7      PRELIMINARY;      PRT;   555 AA.
AC   Q8W0Z7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative branched-chain amino acid aminotransferase
DE   (AT3g05190/T12H1_16).
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Meyers M.C., Banh J., Bowser L.,
RA   Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA   Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA   Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA   Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kim C.J., Chen H., Cheuk R., Meyers M.C., Shinn P., Banh J.,
RA   Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA   Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA   Ecker J.R.;
RT   "Arabidopsis ORF clones.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF462849; AAL58936.1; -.
DR   EMBL; AY090279; AAL90940.1; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
KW   Aminotransferase; Transferase.
SQ   SEQUENCE   555 AA;  62212 MW;  670A17B8D74B5D86 CRC64;

  Query Match             19.3%;  Score 340.5;  DB 10;  Length 555;
  Best Local Similarity   29.5%;  Pred. No. 4.5e-19;
  Matches   88;  Conservative   59;  Mismatches  138;  Indels   13;  Gaps    5;

Qy      47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           :||:  | || | ||:|:||:     |  :    :: | ||:| :|||||  | |  |  |
```

```
Db    263 LAWVGDEILPREMAKVSVFDSVVQGGDSVWEGLRIYKGKIFKLEEHLDRLFDSAKALAFD 322

Qy    107 AGYSKDELAE-ITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL- 164
Db    323 NVPAREEVKEAIFRTLITNGMFDNTHIRLSLTR------GKKVTSGMSPAYNRYGCTLIV 376

Qy    165 ---WAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAIL 221
Db    377 LAEWK-PPVYDNEGGIVLVTATTRRNSPNNLDSKIHHNNLLNNILAKIESNNTNAADAIM 435

Qy    222 LDSDNCVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVFELADQMGIEATLRDVTSRE 280
Db    436 LDKDGYVSETNATNIFMVKKGCVLTPHADYCLPGITRATVMELVVKENFILEERRISLSE 495

Qy    281 LYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIE 338
Db    496 FHTANEVWTTGTMGELSPVVKIDGRVIGDGKVGPVTRTLQNAYKKLTEDSGVPIPTYQ 553


RESULT 7
Q9ASR4
ID   Q9ASR4      PRELIMINARY;      PRT;   559 AA.
AC   Q9ASR4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT5g27410/F21A20_120.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA   Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA   Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
RA   Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K.,
RA   Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF367323; AAK32910.1; -.
DR   HSSP; P00510; 1A3G.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
SQ   SEQUENCE   559 AA;  62896 MW;  6A54870333F54793 CRC64;

  Query Match           18.5%;  Score 326.5;  DB 10;  Length 559;
  Best Local Similarity  28.5%;  Pred. No. 6.1e-18;
  Matches   85;  Conservative   59;  Mismatches  141;  Indels   13;  Gaps    5;

Qy     47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
Db    260 LAWVGDEIVPREMAKVSVFDSVVQGGDSVWEGLRIYKGKVFKLEEHLDRLSDSAKALAFN 319

Qy    107 AGYSKDELAE-ITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL- 164
Db    320 NVPTREEIKEAIFRTLITNGMFDNTHIRLSLTR------GKKVTSGMSPAFNRYGCTLIV 373

Qy    165 ---WAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAIL 221
Db    374 LAEWK-PPVYDNDGGIVLVTATTRRNSPNNLDSKIHHNNLLNNILAKIESNNANVDDAIM 432

Qy    222 LDSDNCVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVFELADQMGIEATLRDVTSRE 280
Db    433 LDKDGFVSETNATNIFMVKKDRVLTPHADYCLPGITRATVMELVVKENFILEERRISLSE 492

Qy    281 LYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIE 338
Db    493 FHTADEVWTTGTMGELSPVVKIDGRVIGEGKVGPVTRRLQNAYKKLTDGSGVPIPTYQ 550


RESULT 8
Q9MAA0
ID   Q9MAA0      PRELIMINARY;      PRT;   490 AA.
AC   Q9MAA0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative branched-chain amino acid aminotransferase.
GN   T12H1.16.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA   Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA   Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT   "Arabidopsis thaliana chromosome III BAC T12H1 genomic sequence.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC009177; AAF27025.1; -.
DR   HSSP; P00510; 1A3G.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
KW   Aminotransferase; Transferase.
SQ   SEQUENCE   490 AA;  54829 MW;  D29B6878564B36BD CRC64;

  Query Match           18.1%;  Score 318.5;  DB 10;  Length 490;
  Best Local Similarity  28.0%;  Pred. No. 2.2e-17;
  Matches   82;  Conservative   51;  Mismatches  123;  Indels   37;  Gaps    3;

Qy     47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
Db    232 LAWVGDEILPREMAKVSVFDSVVQGGDSVWEGLRIYKGKIFKLEEHLDRLFDSAKALAFD 291

Qy    107 AGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWA 166
Db    292 NVPAREEVKEAIFRTLITNVLAE----------------------------------WK 316

Qy    167 FPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDN 226
Db    317 -PPVYDNEGGIVLVTATTRRNSPNNLDSKIHHNNLLNNILAKIESNNTNAADAIMLDKDG 375

Qy    227 CVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVFELADQMGIEATLRDVTSRELYDAD 285
Db    376 YVSETNATNIFMVKKGCVLTPHADYCLPGITRATVMELVVKENFILEERRISLSEFHTAN 435

Qy    286 ELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIE 338
Db    436 EVWTTGTMGELSPVVKIDGRVIGDGKVGPVTRTLQNAYKKLTEDSGVPIPTYQ 488


RESULT 9
Q8ZTE2
ID   Q8ZTE2      PRELIMINARY;      PRT;   303 AA.
AC   Q8ZTE2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Branched-chain amino acid aminotransferase (ilvE).
GN   PAE3297.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
```

```
RX   PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR   EMBL; AE009923; AAL64820.1; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   InterPro; IPR000577; FGGY_kin.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
DR   PROSITE; PS00445; FGGY_KINASES_2; UNKNOWN_1.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   303 AA;  34108 MW;  EB518E66D1BDECA3 CRC64;

  Query Match             17.9%;  Score 315;  DB 17;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 2.1e-17;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :      :   : :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIFEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||        :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
             |   : :   |           ||   : |  ::      |   ||: | :     :
Db     113 RNLEVSLAVIVFP----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | |||| |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| |||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
           : :
Db     281 AELY 284


RESULT 10
Q8ZAB2
ID   Q8ZAB2      PRELIMINARY;      PRT;   308 AA.
AC   Q8ZAB2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Branched-chain amino acid aminotransferase (EC 2.6.1.42).
GN   ILVE OR YPO3899.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414159; CAC93365.1; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   308 AA;  33864 MW;  574CB8C2B6683380 CRC64;

  Query Match             17.1%;  Score 301;  DB 16;  Length 308;
  Best Local Similarity   30.2%;  Pred. No. 2.9e-16;
  Matches   92;  Conservative   46;  Mismatches  141;  Indels   26;  Gaps    9;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IFRLGDHLDRLLDGASKLR 104
           |  || :|  |||: :      :   :    : :    :||  :|: || | |   |
Db       9 WFNGEMVPWAEAKVHVMSHALHYGTSVFEGVRCYESHKGPVVFRHREHMQRLHDSAKIYR 68

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           :    | ||| |  :  :  : |  :::   |  |      |          | | | |:
Db      69 MPVSQSVDELMEACRATLRKNNLTSAYIRPLVFIG-DVGMGVNPPEGYNTDVIIAAFPW- 126

Qy     165 WAFPPAEQI-FGTTAIVPRHVRRAGRNTVDPTIK---NYQWGDLTAASFEAKDRGARTAI 220
           |:  || :  |  |:|      |   ||:    |   ||    |  :  ||:  | :  |
Db     127 GAYLGAEALEQGIDAMVSSW-NRVAPNTIPTAAKAGGNYLSSLLVGS--EARRHGYQEGI 183

Qy     221 LLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKTVFELADQMGIEATLRDVTS 278
           ||    ::|| | |:  |||| | :|  : :|||||||  : :||  ||:|   : ::
Db     184 ALDVHGYLSEGAGENLFEVKDGILFTPPFTSSALPGITRDAIIKLAKDMGLEVREQVLSR 243

Qy     279 RELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF----------WALM 327
             || |||:    ||  :||: |:|| || |: ||:|  |:  |          |  :
Db     244 ESLYLADEVFMSGTAAEITPVRSVDGIQVGIGKRGPVTAKIQQAFFGLFTGETEDKWGWL 303

Qy     328 DEPGP 332
           |:  |
Db     304 DQVNP 308


RESULT 11
Q9HNF8
ID   Q9HNF8      PRELIMINARY;      PRT;   308 AA.
AC   Q9HNF8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Branched-chain amino acid aminotransferase.
GN   ILVE2 OR VNG2122G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005101; AAG20262.1; -.
DR   HSSP; P00510; 1A3G.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01122; ilvE_I; 1.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   308 AA;  34209 MW;  C86A5408B6D93A12 CRC64;

  Query Match             16.9%;  Score 298;  DB 17;  Length 308;
```

```
  Best Local Similarity   27.7%;  Pred. No. 5.1e-16;
  Matches   93;  Conservative   56;  Mismatches  127;  Indels   60;  Gaps   12;

Qy      32 YSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGN----IF 87
           :|: ::||          |::||::  |:|:: :      :    :    |: :    ||
Db       4 FSEMDVDT--------IWMDGEFVDWEDAQVHVLTHALHYGSGIFEGVRVYDTDNGPAIF 55

Qy      88 RLGDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGY---GKRK 144
           |  ||| |  : |    ||  :: ||| : | : :    |   ::      ||   |
Db      56 RWDDHLQRFYESAKPYDLDIEFTPDELTDATVELIQRQDLGSCYIRPLAYYGYESLGVSP 115

Qy     145 GEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQ---- 200
           |:          | | | |  |         |  |:          | ||  :  ::
Db     116 GD-----CPTDVAIAAWP--WG-----AYLGDDAL---------ENGVDVMVSTWRKHAS 154

Qy     201 ------------WGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP- 247
                       : :   |  ||:  |   ||:|: :  |||||| |: :|:||:| :|
Db     155 SQIPTNAKTTGLYVNSMLAGEEARRNGFTEAIVLNKEGNVAEGPGENIFLVRDGELFTPG 214

Qy     248 -SRNALPGITRKTVFELADQMGIEATLRD---VTSRELYDADELMAVTTAGGVTPINSLD 303
            | : | |||| ||  ||:: |   |: |   ::  ||: ||||    :|  ||||  :|
Db     215 LSESILDGITRDTVITLAEERGY--TVHDDVSISRGELHTADELFFTGSAAEVTPIKQVD 272

Qy     304 GEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339
           :|:|  ||:|  ::  |: |: | |   |  .|
Db     273 NVEIGSGTRGPVTEELQTAFFDLV-EAGDREEWFHY 307


RESULT 12
Q9KWZ6
ID   Q9KWZ6      PRELIMINARY;      PRT;   282 AA.
AC   Q9KWZ6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   D-alanine aminotransferase.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COL;
RX   MEDLINE=20031141; PubMed=10566865;
RA   de Lencastre H., Wu S.W., Pinho M.G., Ludovice A.M., Filipe S.,
RA   Gardete S., Sobral R., Gill S., Chung M., Tomasz A.;
RT   "Antibiotic resistance as a stress response: complete sequencing of a
RT   large number of chromosomal loci in Staphylococcus aureus strain COL
RT   that impact on the expression of resistance to methicillin.";
RL   Microb. Drug Resist. 5:163-175(1999).
DR   EMBL; Y14816; CAB82475.1; -.
DR   HSSP; P19938; 1DAA.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; UNKNOWN_1.
KW   Aminotransferase; Transferase.
SQ   SEQUENCE   282 AA;  31894 MW;  6B13D0B038699A0F CRC64;

  Query Match           16.5%;  Score 291;  DB 2;  Length 282;
  Best Local Similarity   27.1%;  Pred. No. 1.7e-15;
  Matches   78;  Conservative   54;  Mismatches  124;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: ||::   |||:|  | |:   |  |    |::| :| : :| :| |  |::: ||
Db       5 FLNGEFVSPSEAKVSYNDRGYVFGDGIYEYIRVYNGKLFTVTEHYERFLRSANEIGLDLN 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           || :|| |:::| | |:|:    : :  ||| :|       ::   :  |   |   :
Db      65 YSVEELIELSRKLVDMNQIETGAIYIQATRGVAERNHSFPTPEVEPAIVAYTKSYDRPYD 124

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGAR 217
           |                    | | |  |::: :|          |:: |  :  |
Db     125 HLEN--------------GVNGV--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAVE 167

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
           :       | ||   |   :||| :   |  |  | ||||  : ::|:   |
Db     168 A--IQHRGETVTEGSSSNAYAIKDGVIYTHPINNYILNGITRIVIKKIAEDYNIPFKEET 225

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |   | :|||::  :|:  |||:  |||| | :|: ||:|  ::: |
Db     226 FTVDFLKNADEVIVSSTSAEVTPVIKLDGEPVNDGKVGPITRQLQEGF 273


RESULT 13
Q99TB4
ID   Q99TB4      PRELIMINARY;      PRT;   282 AA.
AC   Q99TB4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   D-alanine aminotransferase.
GN   SAV1750 OR SA1571.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003363; BAB57912.1; -.
DR   EMBL; AP003135; BAB42839.1; -.
DR   HSSP; P19938; 1DAA.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
DR   TIGRFAMs; TIGR01121; D_amino_aminoT; 1.
DR   PROSITE; PS00770; AA_TRANSFER_CLASS_4; UNKNOWN_1.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   282 AA;  31908 MW;  6B0D215F38699A0F CRC64;

  Query Match           16.4%;  Score 290;  DB 16;  Length 282;
  Best Local Similarity   26.7%;  Pred. No. 2e-15;
  Matches   77;  Conservative   55;  Mismatches  124;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: ||::   |||:|  | |:   |  |    |::| :| : :| :| |  |::: ||
Db       5 FLNGEFVSPSEAKVSYNDRGYVFGDGIYEYIRVYNGKLFTVTEHYERFLRSANEIGLDLN 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           || :|| |:::| | |:|:    : :  ||| :|       ::   :  |   |   :
Db      65 YSVEELIELSRKLVDMNQIETGAIYIQATRGVAERNHSFPTPEVEPAIVAYTKSYDRPYD 124

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGAR 217
           |                    | | |  |::: :|          |:: |  :  |
Db     125 HLEN--------------GVNGV--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAVE 167

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
           :       | ||   |   :||| :   |  |  | ||||  : ::|:   |
Db     168 A--IQHRGETVTEGSSSNAYAIKDGVIYTHPINNYILNGITRIVIKKIAEDYNIPFKEET 225

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |   | :|||::  :|:  |||:  |||| : :|: ||:|  ::: |
Db     226 FTVDFLKNADEVIVSSTSAEVTPVIKLDGEPVNDGKVGPITRQLQEGF 273
```

```
Db    226 FTVDFLKNADEVIVSSTSAEVTPVIKLDGEPINDGKVGPITRQLQEGF 273


RESULT 14
Q8XRW7
ID   Q8XRW7      PRELIMINARY;      PRT;   290 AA.
AC   Q8XRW7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative D-alanine aminotransferase (D-aspartate aminotransferase)
DE   protein (EC 2.6.1.21).
GN   RSP0714 OR RS01735.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OG   Plasmid megaplasmid.
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646080; CAD17865.1; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
KW   Transferase; Aminotransferase; Plasmid; Complete proteome.
SQ   SEQUENCE   290 AA;  31181 MW;  DC054A9EB6098A3A CRC64;

  Query Match             16.4%;  Score 289.5;  DB 16;  Length 290;
  Best Local Similarity   31.4%;  Pred. No. 2.3e-15;
  Matches   90;  Conservative   41;  Mismatches  125;  Indels   31;  Gaps    8;

Qy     47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           : :: |:|:|| || :|| | ||  :|  | |  |  | :     || ||    |:: :|
Db      4 IVFLNGQYVPAAEATVSIMDRGFTFADGIYEVTAVARGKLVDNDAHLARLTRSLSEIGID 63

Qy    107 AGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWA 166
            |:  |   : :  :: : | |  | : |||  :|               : ||   |
Db     64 NPYTDAEWTRVCEALIARNGLEEGVVYMQVTRGAAERD-------------FGIPAQIA 109

Qy    167 FPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWG--DLTA--------ASFEAKDRGA 216
           | |       :||   : | |   |  |: : :|   |: :        |   |   ||
Db    110 -PTAVAFTQVKSIVDSPLARKGAAVV--TVPDLRWKRCDIKSVGLLPQVMAKQVAARAGA 166

Qy    217 RTAILLDSDNCVAEGPGFNVCIVKDGK--LASPSRNA-LPGITRKTVFELADQMGIEATL 273
            | : | |  | ||      |:  |  :  |  || |||||| :|  || : |:
Db    167 SEAWMTDGDR-VTEGASSTAFIITTDKRLITRPLSNAVLPGITRVSVMALAREHGLVLEE 225

Qy    274 RDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIR 320
           |  | :|   | |    : :  | |: |:||  ||:|:|||| |:|
Db    226 RTFTVQEAQQAAEAFFTSASTFVMPVISIDGVPVGDGQPGPLTRALR 272


RESULT 15
Q92Q85
ID   Q92Q85      PRELIMINARY;      PRT;   287 AA.
AC   Q92Q85;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative aminotransferase protein (EC 2.6.1.).
GN   R01463 OR SMC01047.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR   EMBL; AL591787; CAC46042.1; -.
DR   InterPro; IPR001544; Aminotran_4.
DR   Pfam; PF01063; aminotran_4; 1.
DR   ProDom; PD001961; Aminotran_4; 1.
KW   Transferase; Aminotransferase; Complete proteome.
SQ   SEQUENCE   287 AA;  31646 MW;  9062E826F3676257 CRC64;

  Query Match             15.8%;  Score 278;  DB 16;  Length 287;
  Best Local Similarity   26.5%;  Pred. No. 1.9e-14;
  Matches   82;  Conservative   51;  Mismatches  121;  Indels   56;  Gaps    7;

Qy     47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           :|:: | |:   || | : | |:  :|  | |  | || |  |  |||||     :||:
Db      4 IAYVNGTYVRHAEAAIHVEDRGYQFADGVYEVCEVRHGFIIDLTRHLDRLGRSLGELRIG 63

Qy    107 AGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWA 166
            |:  |  : :: :  :::|     | ||||  :|
Db     64 WPMSRAALIHVIREVLRRNRVRNGLFYLQVTRGVARR---------------------- 100

Qy    167 FPPAEQIFGTTAIVPRHVRRAGRNTVDP-------------TIKNYQWGDLTAASF--- 209
            : :|      |  |  | |  ||             |:  :|  :  |
Db    101 ----DHVFPDADTPPSIVVTAKR--TDPGAIARKNAEGISAITVPENRWDRVDIKSVGLL 154

Qy    210 -------EAKDRGARTAILLDSDNCVAEGPGFNVCIV-KDGKLAS-PSRNA-LPGITRKT 259
                 :||: ||: || :| |  | ||   || || ::| | : |: :  | |:|| |
Db    155 PNVLARQQAKEAGAQEAIFVDVDGMVKEGAATNVWIVDREGTLRTRPAESGILRGVTRTT 214

Qy    260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           : ::|  :|:       :  |:  | |:          |: |:||: :||| || :   |
Db    215 LMDVAKPLGLAIEETAFSVEEMLAAREVFITAATSICFPVVSVDGKTIGNGHPGSIAQNI 274

Qy    320 RDRFWALMDE 329
           |: |: : ::
Db    275 RETFFDIAEK 284


Search completed: April 17, 2003, 17:39:32
Job time : 92 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:36:48 ; Search time 15 Seconds
(without alignments)
664.958 Million cell updates/sec

Title: US-10-067-291-1
Perfect score: 1764
Sequence: 1 MTALSDLGTSNLVAVEPGAI.....RDRFWALMDEPGPLIETIEY 339

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1764	100.0	339	4	US-09-527-522-1	Sequence 1, Appli
2	385	21.8	288	4	US-09-173-300-29	Sequence 29, Appl
3	322.5	18.3	310	4	US-09-173-300-31	Sequence 31, Appl
4	315	17.9	303	2	US-08-599-171A-32	Sequence 32, Appl
5	315	17.9	303	2	US-08-646-590B-32	Sequence 32, Appl
6	315	17.9	303	3	US-09-069-226-32	Sequence 32, Appl
7	315	17.9	303	4	US-09-412-184-32	Sequence 32, Appl
8	289	16.4	289	3	US-08-972-902-6	Sequence 6, Appli
9	283	16.0	282	3	US-08-972-902-7	Sequence 7, Appli
10	278.5	15.8	283	3	US-08-972-902-9	Sequence 9, Appli
11	250	14.2	283	3	US-08-972-902-8	Sequence 8, Appli
12	250	14.2	283	4	US-09-466-257A-8	Sequence 8, Appli
13	250	14.2	283	4	US-08-424-797A-3	Sequence 3, Appli
14	246	13.9	283	4	US-09-466-257A-10	Sequence 10, Appl
15	245	13.9	283	4	US-09-134-001C-5100	Sequence 5100, Ap
16	244	13.8	283	1	US-08-723-896-3	Sequence 3, Appli
17	238	13.5	285	4	US-09-173-300-24	Sequence 24, Appl
18	226.5	12.8	255	4	US-09-173-300-33	Sequence 33, Appl
19	222	12.6	297	4	US-09-173-300-26	Sequence 26, Appl
20	165.5	9.4	362	4	US-09-134-001C-5403	Sequence 5403, Ap
21	163.5	9.3	348	4	US-09-173-300-19	Sequence 19, Appl
22	161.5	9.2	243	4	US-09-173-300-22	Sequence 22, Appl
23	154.5	8.8	307	4	US-09-173-300-15	Sequence 15, Appl
24	146	8.3	198	4	US-09-173-300-11	Sequence 11, Appl
25	139.5	7.9	180	4	US-09-173-300-28	Sequence 28, Appl
26	110.5	6.3	115	4	US-09-173-300-13	Sequence 13, Appl
27	102.5	5.8	363	4	US-09-173-300-20	Sequence 20, Appl
28	100	5.7	19	4	US-09-527-522-5	Sequence 5, Appli
29	99.5	5.6	607	2	US-08-472-534-5	Sequence 5, Appli
30	91	5.2	2860	2	US-08-826-267-2	Sequence 2, Appli
31	89	5.0	625	1	US-08-365-981-13	Sequence 13, Appl
32	88.5	5.0	4551	3	US-09-320-878-1	Sequence 1, Appli
33	88.5	5.0	4613	4	US-09-105-537-31	Sequence 31, Appl
34	88.5	5.0	11877	4	US-09-105-537-6	Sequence 6, Appli
35	88	5.0	287	1	US-08-365-981-9	Sequence 9, Appli
36	86.5	4.9	380	2	US-08-846-762-86	Sequence 86, Appl
37	86.5	4.9	491	4	US-09-134-001C-4727	Sequence 4727, Ap
38	85	4.8	413	2	US-08-282-197C-49	Sequence 49, Appl
39	82.5	4.7	303	4	US-09-134-001C-4256	Sequence 4256, Ap
40	81.5	4.6	600	4	US-09-352-159-36	Sequence 36, Appl
41	81.5	4.6	600	4	US-09-352-159-38	Sequence 38, Appl
42	81	4.6	700	1	US-07-720-589-2	Sequence 2, Appli
43	81	4.6	700	2	US-08-785-190-2	Sequence 2, Appli
44	81	4.6	700	4	US-08-235-836C-66	Sequence 66, Appl
45	81	4.6	700	5	PCT-US92-05539-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-527-522-1
; Sequence 1, Application US/09527522
; Patent No. 6413752
; GENERAL INFORMATION:
; APPLICANT: Takashima, Yoshiki
; APPLICANT: Mitsuda, Satoshi
; TITLE OF INVENTION: Protein capable of catalyzing transamination
; TITLE OF INVENTION: stereoselectively, gene encoding said protein
; TITLE OF INVENTION: and use thereof
; FILE REFERENCE: Q58251
; CURRENT APPLICATION NUMBER: US/09/527,522
; CURRENT FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: JP 11/075511
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 11/088634
; PRIOR FILING DATE: 1999-03-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Mycobacterium aurum
; FEATURE:
; OTHER INFORMATION: SC-S423
US-09-527-522-1

Query Match 100.0%; Score 1764; DB 4; Length 339;
Best Local Similarity 100.0%; Pred. No. 3.9e-191;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MTALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MTALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60

Qy      61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120

Qy     121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180

Qy     181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240

Qy     241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300
```

```
Qy    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339
          |||||||||||||||||||||||||||||||||||||||
Db    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339


RESULT 2
US-09-173-300-29
; Sequence 29, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT: Hitz, William D.
;  APPLICANT: Kinney, Anthony J.
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/09/173,300
;  CURRENT FILING DATE: 1998-10-15
;  EARLIER APPLICATION NUMBER: 60/063,423
;  EARLIER FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 29
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Methanococcus jannaschii
US-09-173-300-29

  Query Match           21.8%;  Score 385;  DB 4;  Length 288;
  Best Local Similarity 31.9%;  Pred. No. 4.2e-35;
  Matches   90;  Conservative   58;  Mismatches  116;  Indels   18;  Gaps    5;

Qy     49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
          :: |:::  ::||:|:|| |  : |  :     : | :| | :|:||| | |  | :|
Db      4 YLNGKFVDEKDAKVSVFDHGLLYGDGVFEGIRAYDGVVFMLKEHIDRLYDSAKSLCIDIP 63

Qy    109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYG-----KRKGEKDLSKLTHQVYIYAIPY 163
           :|:|: ::  : : :: ||:::: | |||| |      ||  |         ::  |||
Db     64 LTKEEMIDVVLETLRVNNLRDAYIRLVVTRGVGDLGLDPRKCGKP------TIFCIAIP- 116

Qy    164 LWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
              ||     |  ||    |||   : ::| :|:  : :   |  :|   |   | |||
Db    117 ---MPPLLGEDGIRAITV-SVRRLPVDVLNPAVKSLNYLNSVLAKIQANYAGVDEAFLLD 172

Qy    224 SDNCVAEGPGFNVCIVKDGKLASPS--RNALPGITRKTVFELADQMGIEATLRDVTSREL 281
              | || | |: |||:| | :|   :: | ||||  | :|| : |||     :|  :|
Db    173 DKGFVVEGTGDNIFIVKNGVLKTPPVYQSILKGITRDVVIKLAKEEGIEVVEEPLTLHDL 232

Qy    282 YDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
          | ||||    ||  : |:  :||  : | : | :|  ::::|
Db    233 YTADELFITGTAAEIVPVFEIDGRVINNKQVGEITKKLKEKF 274


RESULT 3
US-09-173-300-31
; Sequence 31, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT: Hitz, William D.
;  APPLICANT: Kinney, Anthony J.
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/09/173,300
;  CURRENT FILING DATE: 1998-10-15
;  EARLIER APPLICATION NUMBER: 60/063,423
;  EARLIER FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
;
; SEQ ID NO 31
;   LENGTH: 310
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-173-300-31

  Query Match           18.3%;  Score 322.5;  DB 4;  Length 310;
  Best Local Similarity 29.0%;  Pred. No. 5.7e-28;
  Matches   82;  Conservative   54;  Mismatches  134;  Indels   13;  Gaps    5;

Qy     47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
          : |:  | ||   ||:|:|||:     |  :    :: | :|:| :||||| | |  :
Db     18 LVWVGDELLPRNSAKVSVFDSVVQGGDAVWEGLRIYDGKVFKLDEHLDRLFDSAKAMAFS 77

Qy    107 AGYSKDELAE-ITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL- 164
             ::| : : | |  ::      : : ||:||      |:|  | ::   :|    :
Db     78 NVPTRDWIKDAIFKTLIANGMFNNAHIRLTLTR------GKKVTSGMSPAFNLYGCALIV 131

Qy    165 ---WAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAIL 221
             |  |  :   |   :|    ||   |::|| | :    :   |  |     |  ||:
Db    132 LAEWKPPVYDNSHG-IKLVTATTRRNSPNSIDPKIHHNNLINNILAKIEGNLAQAEDAIM 190

Qy    222 LDSDNCVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVFELADQMGIEATLRDVTSRE 280
          || |  |:|    |: :|| | : :|  :  ||||| || :|  :        | ::  |
Db    191 LDKDGFVSETNATNIFMVKKGIVLTPHADYCLPGITRATVMDLVVKENFVLHERRISLSE 250

Qy    281 LYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
          : |||:    | | :||:  :||  :|:|: ||:|  |:  :
Db    251 FHAADEVWTTGTMGEITPVVMIDGREIGDGKIGPVTRQIQKAY 293


RESULT 4
US-08-599-171A-32
; Sequence 32, Application US/08599171A
; Patent No. 5814473
;  GENERAL INFORMATION:
;    APPLICANT:  WARREN, Patrick V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/599,171A
;      FILING DATE:  Concurrently
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HERRON, CHARLES J.
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-38
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 AMINO ACIDS
;      TYPE:  AMINO ACID
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
```

```
US-08-599-171A-32

  Query Match             17.9%;  Score 315;  DB 2;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 3.9e-27;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :      :    :  :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIFEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||       :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
             |   : :   |                ||   : |  ::      |   ||: | :      :
Db     113 RNLEVSLAVIVFP----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | |||| |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| |||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
           : :
Db     281 AELY 284


RESULT 5
US-08-646-590B-32
; Sequence 32, Application US/08646590B
; Patent No. 5962283
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Patrick V.
;    APPLICANT:  Swanson, Ronald V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,590B
;      FILING DATE:  08-May-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/599,171
;      FILING DATE:  09-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/01094
;      FILING DATE:  21-January-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/0170O1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-08-646-590B-32

  Query Match             17.9%;  Score 315;  DB 2;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 3.9e-27;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :      :    :  :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIFEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||       :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
             |   : :   |                ||   : |  ::      |   ||: | :      :
Db     113 RNLEVSLAVIVFP----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | |||| |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| |||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
           : :
Db     281 AELY 284


RESULT 6
US-09-069-226-32
; Sequence 32, Application US/09069226
; Patent No. 6013509
;  GENERAL INFORMATION:
;    APPLICANT:  WARREN, Patrick V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/069,226
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/599,171
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HERRON, CHARLES J.
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-38
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 AMINO ACIDS
```

```
;      TYPE:  AMINO ACID
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-09-069-226-32

  Query Match            17.9%;  Score 315;  DB 3;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 3.9e-27;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :      :    :  :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIPEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||       :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
             |   : :   |           ||   : |  ::      |   ||: | :      :
Db     113 RNLEVSLAVIVFP-----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | |||| |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| |||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
           : :
Db     281 AELY 284


RESULT 7
US-09-412-184-32
; Sequence 32, Application US/09412184
; Patent No. 6268188
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Patrick V.
;    APPLICANT:  Swanson, Ronald V.
;    TITLE OF INVENTION:  TRANSAMINASES AND AMINOTRANSFERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/412,184
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,590
;      FILING DATE:  08-May-1996
;      APPLICATION NUMBER:  08/599,171
;      FILING DATE:  09-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/01094
;      FILING DATE:  21-January-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haile, Ph.D., Lisa A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/017001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619/678-5070
;      TELEFAX:  619/678-5099
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  303 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-09-412-184-32

  Query Match            17.9%;  Score 315;  DB 4;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 3.9e-27;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :      :    :  :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIFEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||       :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
             |   : :   |           ||   : |  ::      |   ||: | :      :
Db     113 RNLEVSLAVIVFP-----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | |||| |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| |||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
           : :
Db     281 AELY 284


RESULT 8
US-08-972-902-6
; Sequence 6, Application US/08972902
; Patent No. 6099848
;  GENERAL INFORMATION:
;    APPLICANT:  Frankel, Fred R.
;    APPLICANT:  Portnoy, Daniel A.
;    TITLE OF INVENTION:  BACTERIAL VACCINE VECTOR AND METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  One Commerce Square, 2005 Market Street, 22nd
;      STREET:  Fl.
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/972,902
;      FILING DATE:  18-NOV-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Doyle-Leary, Kathryn
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9596-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
```

```
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;    INFORMATION FOR SEQ ID NO:  6:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  289 amino acids
;        TYPE:  amino acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-972-902-6

  Query Match             16.4%;  Score 289;  DB 3;  Length 289;
  Best Local Similarity   28.9%;  Pred. No. 3.2e-24;
  Matches   83;  Conservative   51;  Mismatches  123;  Indels   30;  Gaps    6;

Qy      50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109
           :    :  |:| : | | |:   |  | |  :::|  |   :|:|||   |:|: |   |
Db       5 VNNHLVEREDATVDIEDRGYQFGDGVYEVVRLYNGKFFTYNEHIDRLYASAAKIDLVIPY 64

Qy     110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPP 169
           ||:|| |: :| |: : :    | | ||||    :               : ||    ||
Db      65 SKEELRELLEKLVAENNINTGNVYLQVTRGVQNPRN----------HVIPD--DFPL 109

Qy     170 AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGART 218
              :      |||: |:      | :: :|          |:: | : :|  : |
Db     110 EGVLTAAAREVPRNERQFVEGGTAITEEDVRWLRCDIKSLNLLGNILAKN-KAHQQNALE 168

Qy     219 AILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALPGITRKTVFELADQMGIEATLRDV 276
           |||   :  | |    || |:||| | : : :   | ||||: : ::| : ||    |
Db     169 AILHRGEQ-VTECSASNVSIIKDGVLWTHAADNLILNGITRQVIIDVAKKNGIPVKEADF 227

Qy     277 TSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |  :| :|||:   :|   :|||  :||  | :|: ||:|  :   |
Db     228 TLTDLREADEVFISSTTIEITPITHIDGVQVADGKRGPITAQLHQYF 274


RESULT 9
US-08-972-902-7
; Sequence 7, Application US/08972902
; Patent No. 6099848
;  GENERAL INFORMATION:
;    APPLICANT:  Frankel, Fred R.
;    APPLICANT:  Portnoy, Daniel A.
;    TITLE OF INVENTION:  BACTERIAL VACCINE VECTOR AND METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  One Commerce Square, 2005 Market Street, 22nd
;      STREET:  Fl.
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/972,902
;      FILING DATE:  18-NOV-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Doyle-Leary, Kathryn
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9596-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  282 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-972-902-7

  Query Match             16.0%;  Score 283;  DB 3;  Length 282;
  Best Local Similarity   25.3%;  Pred. No. 1.5e-23;
  Matches   73;  Conservative   59;  Mismatches  124;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :| ||::   |||:|  | |:   |  |     : | :| : :| :| :  ||:::|| |
Db       5 FINGEFIDQNEAKVSYEDRGYVFGDGIYEYIRAYDGKLFTVTEHFERFIRSASEIQLDLG 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           |: :|| :: :: : :: ::   : :  |||   |       ::   :  :|  |   :
Db      65 YTVEELIDVVRELLKVNNIQNGGIYIQATRGVAPRNHSFPTPEVKPVIMAFAKSYDRPYD 124

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGAR 217
             |               | |    |::: :|          |:: |  :  |
Db     125 DLEN--------------GINAA--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAGE 167

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
           :       | ||   ||  :||| :   |  |  | ||||| :  :::   |
Db     168 A--IQHRGETVTEGASSNVYAIKDGAIYTHPVNNYILNGITRKVIKWISEDEDIPFKEET 225

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
             |  | :|||::  :|:  |||:  :||| ||:|: ||:|  :::
Db     226 FTVEFLKNADEVIVSSTSAEVTPVVKIDGEQVGDGKVGPVTRQLQEGF 273


RESULT 10
US-08-972-902-9
; Sequence 9, Application US/08972902
; Patent No. 6099848
;  GENERAL INFORMATION:
;    APPLICANT:  Frankel, Fred R.
;    APPLICANT:  Portnoy, Daniel A.
;    TITLE OF INVENTION:  BACTERIAL VACCINE VECTOR AND METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  One Commerce Square, 2005 Market Street, 22nd
;      STREET:  Fl.
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/972,902
;      FILING DATE:  18-NOV-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Doyle-Leary, Kathryn
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9596-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  283 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
US-08-972-902-9

  Query Match             15.8%;  Score 278.5;  DB 3;  Length 283;
  Best Local Similarity   27.6%;  Pred. No. 4.7e-23;
  Matches   83;  Conservative   50;  Mismatches  131;  Indels   37;  Gaps    9;

Qy      45 GGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLR 104
           |   |    : :  || ||   | |:   |  | |  |::| :| : :|:|||   | |:|
Db       2 GYTLW-NDQIVKDEEVKIDKEDRGYQFGDGVYEVVKVYNGEMFTVNEHIDRLYASAEKIR 60

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           :   |:||:  ::  : |  ::|    :   ||||   |          ||
Db      61 ITIPYTKDKFHQLLHELVEKNELNTGHIYFQVTRGTSPR---------AHQ--------- 102

Qy     165 WAFPP---AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW--GDLTA--------ASFEA 211
             ||        | | |   || :     :     ::: :|   |: :        |  ||
Db     103 --FPENTVKPVIIGYTKENPRPLENLEKGVKATFVEDIRWLRCDIKSLNLLGAVLAKQEA 160

Qy     212 KDRGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGI 269
           ::|   || |  :| | ||   ||  :||| |    |: |  | ||||  |   |::: :
Db     161 HEKGCYEAI-LHRNNTVTEGSSSNVFGIKDGILYTHPANNMILKGITRDVVIACANEINM 219

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
             |: |    |||   :|   :||:  :||: : :|: |  |  :: :|   : :
Db     220 PVKEIPFTTHEALKMDELFVTSTTSEITPVIEIDGKLIRDGKVGEWTRKLQKQFETKIPK 279

Qy     330 P 330
           |
Db     280 P 280


RESULT 11
US-08-972-902-8
; Sequence 8, Application US/08972902
; Patent No. 6099848
;  GENERAL INFORMATION:
;    APPLICANT:  Frankel, Fred R.
;    APPLICANT:  Portnoy, Daniel A.
;    TITLE OF INVENTION:  BACTERIAL VACCINE VECTOR AND METHODS OF
;    TITLE OF INVENTION:  USE THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PANITCH SCHWARZE JACOBS & NADEL, P.C.
;      STREET:  One Commerce Square, 2005 Market Street, 22nd
;      STREET:  Fl.
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103-7086
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/972,902
;      FILING DATE:  18-NOV-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Doyle-Leary, Kathryn
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  9596-51
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-965-1284
;      TELEFAX:  215-567-2991
;      TELEX:  831-494
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  283 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear

;    MOLECULE TYPE:  protein
US-08-972-902-8

  Query Match             14.2%;  Score 250;  DB 3;  Length 283;
  Best Local Similarity   26.9%;  Pred. No. 8e-20;
  Matches   79;  Conservative   49;  Mismatches  110;  Indels   56;  Gaps    9;

Qy      58 EEAKISIF--DTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELA 115
           ||  |:|   | |:   |  | |  |::|::|   :|:||    | |:||   |:|| |
Db      12 EEGSITISPEDRGYQFGDGIYEVIKVYNGHMFTAQEHIDRFYASAEKIRLVIPYTKDVLH 71

Qy     116 EITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFG 175
           ::    :  : |    |   :|||   |                        || |
Db      72 KLLHDLIEKNNLNTGHVYFQITRGTTSRN-------------------HIFPDA----S 107

Qy     176 TTAIVPRHVRRAGRNTVDPTIKNYQWG-------------DLTA--------ASFEAKD 213
             |::  :|:     | : :|:|:: |              |: :        |  || :
Db     108 VPAVLTGNVK-----TGERSIENFEKGVKATLVEDVRWLRCDIKSLNLLGAVLAKQEASE 162

Qy     214 RGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEA 271
           :|   |||   |  : |    ||  :|||||   |: |  | ||||: : : | :: :
Db     163 KGCYEAILHRGD-IITECSSANVYGIKDGKLYTHPANNYILNGITRQVILKCAAEINLPV 221

Qy     272 TLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWA 325
             :|  :|   ||::  : :  |||:  :||: :| | ||  |  ::  | |
Db     222 IEEPMTKGDLLTMDEIIVSSVSSEVTPVIDVDGQQIGAGVPGEWTRKLQKAFEA 275


RESULT 12
US-09-466-257A-8
; Sequence 8, Application US/09466257A
; Patent No. 6337190
; GENERAL INFORMATION:
;  APPLICANT:  Hwang, Tzann-Shun
;  APPLICANT:  Wu, Szu-Pei
;  APPLICANT:  Chou, Hsin-Hua
;  APPLICANT:  Chen, Hwa-Yi
;  APPLICANT:  Lin, Lung-Shen
;  APPLICANT:  Tsai, Hsin
;  APPLICANT:  Chang, Edward
;  TITLE OF INVENTION:  A No. 6337190el D-Amino Acid Aminotransferase For
;  TITLE OF INVENTION:  Simultaneously Producing
;  TITLE OF INVENTION:  Glutaryl-7-Aminocephalosporanic Acid And D-Amino Acid
;  FILE REFERENCE:  1476-4003
;  CURRENT APPLICATION NUMBER:  US/09/466,257A
;  CURRENT FILING DATE:  1999-12-17
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH:  283
;   TYPE:  PRT
;   ORGANISM:  Bacillus sphaericus
US-09-466-257A-8

  Query Match             14.2%;  Score 250;  DB 4;  Length 283;
  Best Local Similarity   26.9%;  Pred. No. 8e-20;
  Matches   79;  Conservative   49;  Mismatches  110;  Indels   56;  Gaps    9;

Qy      58 EEAKISIF--DTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELA 115
           ||  |:|   | |:   |  | |  |::|::|   :|:||    | |:||   |:|| |
Db      12 EEGSITISPEDRGYQFGDGIYEVIKVYNGHMFTAQEHIDRFYASAEKIRLVIPYTKDVLH 71

Qy     116 EITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFG 175
           ::    :  : |    |   :|||   |                        || |
Db      72 KLLHDLIEKNNLNTGHVYFQITRGTTSRN-------------------HIFPDA----S 107

Qy     176 TTAIVPRHVRRAGRNTVDPTIKNYQWG-------------DLTA--------ASFEAKD 213
             |::  :|:     | : :|:|:: |              |: :        |  || :
Db     108 VPAVLTGNVK-----TGERSIENFEKGVKATLVEDVRWLRCDIKSLNLLGAVLAKQEASE 162

Qy     214 RGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEA 271
           :|   |||   |  : |    ||  :|||||   |: |  | ||||: : : | :: :
Db     163 KGCYEAILHRGD-IITECSSANVYGIKDGKLYTHPANNYILNGITRQVILKCAAEINLPV 221
```

```
Qy    272 TLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWA 325
Db    222 IEEPMTKGDLLTMDEIIVSSVSSEVTPVIDVDGQQIGAGVPGEWTRKLQKAFEA 275


RESULT 13
US-08-424-797A-3
; Sequence 3, Application US/08424797A
; Patent No. 6358714
;  GENERAL INFORMATION:
;    APPLICANT:  Fotheringham, Ian G.
;    APPLICANT:  Taylor, Paul P.
;    APPLICANT:  Yoshida, Roberta K.
;    TITLE OF INVENTION:  Materials and Methods for the
;    TITLE OF INVENTION:  Production of D-phenylalanine
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive, 6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/424,797A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  27127/32582
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  283 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-424-797A-3

  Query Match             14.2%;  Score 250;  DB 4;  Length 283;
  Best Local Similarity   26.9%;  Pred. No. 8e-20;
  Matches   79;  Conservative   49;  Mismatches  110;  Indels   56;  Gaps    9;

Qy     58 EEAKISIF--DTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELA 115
Db     12 EEGSITISPEDRGYQFGDGIYEVIKVYNGHMFTAQEHIDRFYASAEKIRLVIPYTKDVLH 71

Qy    116 EITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFG 175
Db     72 KLLHDLIEKNNLNTGHVYFQITRGTTSRN-------------------HIFPDA----S 107

Qy    176 TTAIVPRHVRRAGRNTVDPTIKNYQWG--------------DLTA--------ASFEAKD 213
Db    108 VPAVLTGNVK-----TGERSIENFEKGVKATLVEDVRWLRCDIKSLNLLGAVLAKQEASE 162

Qy    214 RGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEA 271
Db    163 KGCYEAILHRGD-IITECSSANVYGIKDGKLYTHPANNYILNGITRQVILKCAAEINLPV 221

Qy    272 TLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWA 325
Db    222 IEEPMTKGDLLTMDEIIVSSVSSEVTPVIDVDGQQIGAGVPGEWTRKLQKAFEA 275


RESULT 14
US-09-466-257A-10
; Sequence 10, Application US/09466257A
; Patent No. 6337190
; GENERAL INFORMATION:
;  APPLICANT:  Hwang, Tzann-Shun
;  APPLICANT:  Wu, Szu-Pei
;  APPLICANT:  Chou, Hsin-Hua
;  APPLICANT:  Chen, Hwa-Yi
;  APPLICANT:  Lin, Lung-Shen
;  APPLICANT:  Tsai, Hsin
;  APPLICANT:  Chang, Edward
;  TITLE OF INVENTION:  A No. 6337190el D-Amino Acid Aminotransferase For
;  TITLE OF INVENTION:  Simultaneously Producing
;  TITLE OF INVENTION:  Glutaryl-7-Aminocephalosporanic Acid And D-Amino Acid
;  FILE REFERENCE:  1476-4003
;  CURRENT APPLICATION NUMBER: US/09/466,257A
;  CURRENT FILING DATE:  1999-12-17
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 283
;   TYPE: PRT
;   ORGANISM: Bacillus sphaericus
US-09-466-257A-10

  Query Match             13.9%;  Score 246;  DB 4;  Length 283;
  Best Local Similarity   26.7%;  Pred. No. 2.3e-19;
  Matches   77;  Conservative   48;  Mismatches  109;  Indels   54;  Gaps    8;

Qy     62 ISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKKC 121
Db     18 ISPEDRGYQFGDGIYEVIKVYNGHMFTAQEHIDRFYASAEKIRLVIPYTKDVLHKLLHDL 77

Qy    122 VSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVP 181
Db     78 IEKNNLNTGHVYFQITRGTTSRN-------------------HIFPDA----SVPAVLT 113

Qy    182 RHVRRAGRNTVDPTIKNYQWG--------------DLTA--------ASFEAKDRGARTA 219
Db    114 GNVK-----TGERSIENFEKGVKATLVEDVRWLRCDIKSLNLLGAVLAKQEASEKGCYEA 168

Qy    220 ILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRDVT 277
Db    169 ILHRGD-IITECSSANVYGIKDGKLYTHPANNYILNGITRQVILKCAAEINLPVIEEPMT 227

Qy    278 SRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWA 325
Db    228 KGDLLTMDEIIVSSVSSEVTPVIDVDGQQIGAGVPGEWTRKLQKAFEA 275


RESULT 15
US-09-134-001C-5100
; Sequence 5100, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5100
;   LENGTH: 283
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5100
```

```
  Query Match             13.9%;  Score 245;  DB 4;  Length 283;
  Best Local Similarity   25.3%;  Pred. No. 2.9e-19;
  Matches   73;  Conservative   51;  Mismatches  132;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :| ||::  |:||:|  | |:   |  |     : | :| : :| :| |  | :: ||
Db       6 FINGEFVNEEDAKVSYEDRGYVFGDGIYEYIRAYDGKLFTVKEHFERFLRSAEEIGLDLN 65

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           |: :|| |: :: :  : :    : :  |||  |       :   :  :   |   :
Db      66 YTIEELIELVRRLLKENNVVNGGIYIQATRGAAPRNHSFPTPPVKPVIMAFTKSYDRPYE 125

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGAR 217
             ||  |  ||              | :: :|          |:: |  :  |   |
Db     126 ELEQ--GVYAI--------------TTEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAAE 168

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
              :   : | ||   ||  :||| :   |  |  | ||||: :  :|:   |
Db     169 A--IQHRGDIVTEGASSNVYAIKDGVIYTHPVNNFILNGITRRVIKWIAEDEQIPFKEET 226

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |   |  |||::  :|:  | ||  :||| | :|: | :|  ::  |
Db     227 FTVEFLKSADEVIISSTSAEVMPITKIDGENVQDGQVGTITRQLQQGF 274


Search completed: April 17, 2003, 17:40:20
Job time : 20 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:39:37 ; Search time 20 Seconds
(without alignments)
1281.655 Million cell updates/sec

Title: US-10-067-291-1
Perfect score: 1764
Sequence: 1 MTALSDLGTSNLVAVEPGAI.....RDRFWALMDEPGPLIETIEY 339

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1764	100.0	339	12	US-10-067-291-1	Sequence 1, Appli
2	847	48.0	324	9	US-10-112-706-1	Sequence 1, Appli
3	847	48.0	325	9	US-10-112-706-2	Sequence 2, Appli
4	847	48.0	329	9	US-10-112-706-3	Sequence 3, Appli
5	847	48.0	330	9	US-10-112-706-4	Sequence 4, Appli
6	385	21.8	288	12	US-10-027-450-29	Sequence 29, Appl
7	322.5	18.3	310	12	US-10-027-450-31	Sequence 31, Appl
8	315	17.9	303	9	US-10-060-432-32	Sequence 32, Appl
9	315	17.9	303	10	US-09-905-173-32	Sequence 32, Appl
10	289	16.4	289	12	US-10-136-253-6	Sequence 6, Appli
11	283	16.0	282	12	US-10-136-253-7	Sequence 7, Appli
12	278.5	15.8	283	12	US-10-136-253-9	Sequence 9, Appli
13	250	14.2	283	12	US-10-136-253-8	Sequence 8, Appli
14	238	13.5	285	12	US-10-027-450-24	Sequence 24, Appl
15	226.5	12.8	255	12	US-10-027-450-33	Sequence 33, Appl
16	222	12.6	297	12	US-10-027-450-26	Sequence 26, Appl
17	169.5	9.6	179	10	US-09-734-017A-42	Sequence 42, Appl
18	163.5	9.3	348	12	US-10-027-450-19	Sequence 19, Appl
19	161.5	9.2	243	12	US-10-027-450-22	Sequence 22, Appl
20	154.5	8.8	307	12	US-10-027-450-15	Sequence 15, Appl
21	146	8.3	198	12	US-10-027-450-11	Sequence 11, Appl
22	139.5	7.9	180	12	US-10-027-450-28	Sequence 28, Appl
23	137	7.8	131	1	US-08-781-986A-5224	Sequence 5224, Ap
24	133.5	7.6	379	9	US-09-738-626-5914	Sequence 5914, Ap
25	122	6.9	314	9	US-09-738-626-6338	Sequence 6338, Ap
26	110.5	6.3	115	12	US-10-027-450-13	Sequence 13, Appl
27	108	6.1	440	10	US-09-734-569-178	Sequence 178, App
28	102.5	5.8	363	12	US-10-027-450-20	Sequence 20, Appl
29	100.5	5.7	410	9	US-09-746-660A-12	Sequence 12, Appl
30	100.5	5.7	456	9	US-09-746-660A-10	Sequence 10, Appl
31	100	5.7	19	12	US-10-067-291-5	Sequence 5, Appli
32	99.5	5.6	607	10	US-09-815-242-13379	Sequence 13379, A
33	99.5	5.6	607	10	US-09-815-242-13682	Sequence 13682, A
34	96	5.4	438	9	US-09-738-626-6181	Sequence 6181, Ap
35	93	5.3	782	9	US-09-813-408-29	Sequence 29, Appl
36	91	5.2	721	12	US-10-025-187-2	Sequence 2, Appli
37	89	5.0	554	9	US-09-738-626-6289	Sequence 6289, Ap
38	88.5	5.0	664	10	US-09-815-242-13820	Sequence 13820, A
39	88.5	5.0	4613	9	US-09-860-846-31	Sequence 31, Appl
40	88.5	5.0	4613	10	US-09-861-289-31	Sequence 31, Appl
41	88.5	5.0	11877	9	US-09-860-846-6	Sequence 6, Appli
42	88.5	5.0	11877	10	US-09-861-289-6	Sequence 6, Appli
43	85.5	4.8	996	9	US-09-738-626-3847	Sequence 3847, Ap
44	85	4.8	1047	10	US-09-866-582-37	Sequence 37, Appl
45	84.5	4.8	620	10	US-09-815-242-11472	Sequence 11472, A

ALIGNMENTS

RESULT 1
US-10-067-291-1
; Sequence 1, Application US/10067291
; Patent No. US20020127664A1
; GENERAL INFORMATION:
; APPLICANT: Takashima, Yoshiki
; APPLICANT: Mitsuda, Satoshi
; TITLE OF INVENTION: Protein capable of catalyzing transamination
; TITLE OF INVENTION: stereoselectively, gene encoding said protein
; TITLE OF INVENTION: and use thereof
; FILE REFERENCE: Q58251
; CURRENT APPLICATION NUMBER: US/10/067,291
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 09/527,522
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: JP 11/075511
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 11/088634
; PRIOR FILING DATE: 1999-03-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Mycobacterium aurum
; FEATURE:
; OTHER INFORMATION: SC-S423
US-10-067-291-1

Query Match 100.0%; Score 1764; DB 12; Length 339;
Best Local Similarity 100.0%; Pred. No. 2.6e-159;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MTALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MTALSDLGTSNLVAVEPGAIREDTPAGSVIQYSDYELDTSSPFAGGVAWIEGEYLPAEEA 60

Qy      61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 KISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELAEITKK 120

Qy     121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFGTTAIV 180

Qy    181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVK 240

Qy    241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DGKLASPSRNALPGITRKTVFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPIN 300

Qy    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339
          |||||||||||||||||||||||||||||||||||||||
Db    301 SLDGEAVGNGEPGPLTVAIRDRFWALMDEPGPLIETIEY 339


RESULT 2
US-10-112-706-1
; Sequence 1, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
;  APPLICANT: YAMADA, Yukio
;  APPLICANT:  IWASAKI, Akira
;  APPLICANT:  KIZAKI, No. US20020192786A1iyuki
;  APPLICANT:  MATSUMOTO, Keiji
;  APPLICANT:  IKENAKA, Yasuhiro
;  APPLICANT:  OGURA, Masahiro
;  APPLICANT:  HASEGAWA, Junzo
;  TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
;  FILE REFERENCE: 1422-398P
;  CURRENT APPLICATION NUMBER: US/10/112,706
;  CURRENT FILING DATE:  2002-04-02
;  PRIOR APPLICATION NUMBER: 09/403,493
;  PRIOR FILING DATE: 1999-10-22
;  PRIOR APPLICATION NUMBER: JP 9-121732
;  PRIOR FILING DATE: 1997-04-23
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 324
;   TYPE: PRT
;   ORGANISM: Arthrobacter sp.
US-10-112-706-1

  Query Match             48.0%;  Score 847;  DB 9;  Length 324;
  Best Local Similarity   51.0%;  Pred. No. 2.6e-72;
  Matches  158;  Conservative   56;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      30 IQYSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRL 89
           | ||||||| ::| ||| |||| ::|  ||:|||| |: |||:|||| |||:|| |||
Db      15 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRL 74

Qy      90 GDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDL 149
            ||::||   |  :|:    ::||: ||  : |: ::|||:||::::|||    ||:|:
Db      75 DDHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITRGYSSTPGERDI 134

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASF 209
           :|   |||:||:|| |  |       |  |:| : |||  |:::|| :||:|||||  |
Db     135 TKHRPQVYMYAVPYQWIVPFDRIRDGVHAMVAQSVRRTPRSSIDPQVKNFQWGDLIRAVQ 194

Qy     210 EAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGI 269
           |  |||    :||| |  :||| |||| ::||| : || | ||||||||| |:|: :|
Db     195 ETHDRGFEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTVLEIAESLGH 254

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
           || | |:|  || ||||::  |||||| |  |:||  : :| |||:| :|  |:| |  |
Db     255 EAILADITLAELLDADEVLGCTTAGGVWPFVSVDGNPISDGVPGPITQSIIRRYWELNVE 314

Qy     330 PGPLIETIEY 339
             |:  ::|
Db     315 SSSLLTPVQY 324


RESULT 3
US-10-112-706-2
; Sequence 2, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
;  APPLICANT: YAMADA, Yukio
;  APPLICANT:  IWASAKI, Akira
;  APPLICANT:  KIZAKI, No. US20020192786A1iyuki
;  APPLICANT:  MATSUMOTO, Keiji
;  APPLICANT:  IKENAKA, Yasuhiro
;  APPLICANT:  OGURA, Masahiro
;  APPLICANT:  HASEGAWA, Junzo
;  TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
;  FILE REFERENCE: 1422-398P
;  CURRENT APPLICATION NUMBER: US/10/112,706
;  CURRENT FILING DATE:  2002-04-02
;  PRIOR APPLICATION NUMBER: 09/403,493
;  PRIOR FILING DATE: 1999-10-22
;  PRIOR APPLICATION NUMBER: JP 9-121732
;  PRIOR FILING DATE: 1997-04-23
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 325
;   TYPE: PRT
;   ORGANISM: Arthrobacter sp.
US-10-112-706-2

  Query Match             48.0%;  Score 847;  DB 9;  Length 325;
  Best Local Similarity   51.0%;  Pred. No. 2.6e-72;
  Matches  158;  Conservative   56;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      30 IQYSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRL 89
           | ||||||| ::| ||| |||| ::|  ||:|||| |: |||:|||| |||:|| |||
Db      16 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRL 75

Qy      90 GDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDL 149
            ||::||   |  :|:    ::||: ||  : |: ::|||:||::::|||    ||:|:
Db      76 DDHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITRGYSSTPGERDI 135

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASF 209
           :|   |||:||:|| |  |       |  |:| : |||  |:::|| :||:|||||  |
Db     136 TKHRPQVYMYAVPYQWIVPFDRIRDGVHAMVAQSVRRTPRSSIDPQVKNFQWGDLIRAVQ 195

Qy     210 EAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGI 269
           |  |||    :||| |  :||| |||| ::||| : || | ||||||||| |:|: :|
Db     196 ETHDRGFEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTVLEIAESLGH 255

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
           || | |:|  || ||||::  |||||| |  |:||  : :| |||:| :|  |:| |  |
Db     256 EAILADITLAELLDADEVLGCTTAGGVWPFVSVDGNPISDGVPGPITQSIIRRYWELNVE 315

Qy     330 PGPLIETIEY 339
             |:  ::|
Db     316 SSSLLTPVQY 325


RESULT 4
US-10-112-706-3
; Sequence 3, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
;  APPLICANT: YAMADA, Yukio
;  APPLICANT:  IWASAKI, Akira
;  APPLICANT:  KIZAKI, No. US20020192786A1iyuki
;  APPLICANT:  MATSUMOTO, Keiji
;  APPLICANT:  IKENAKA, Yasuhiro
;  APPLICANT:  OGURA, Masahiro
;  APPLICANT:  HASEGAWA, Junzo
;  TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
;  FILE REFERENCE: 1422-398P
;  CURRENT APPLICATION NUMBER: US/10/112,706
;  CURRENT FILING DATE:  2002-04-02
```

```
;  PRIOR APPLICATION NUMBER: 09/403,493
;  PRIOR FILING DATE: 1999-10-22
;  PRIOR APPLICATION NUMBER: JP 9-121732
;  PRIOR FILING DATE: 1997-04-23
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
;   LENGTH: 329
;   TYPE: PRT
;   ORGANISM: Arthrobacter sp.
US-10-112-706-3

  Query Match             48.0%;  Score 847;  DB 9;  Length 329;
  Best Local Similarity   51.0%;  Pred. No. 2.6e-72;  .
  Matches  158;  Conservative   56;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      30 IQYSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRL 89
           | ||||||| ::| ||| ||||| ::|  ||:||||| |: |||:|||| |||:|| |||
Db      20 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRL 79

Qy      90 GDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDL 149
             ||::||   |  :|:    ::||: ||  : |: ::||||:||::::|||    ||:|:
Db      80 DDHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITRGYSSTPGERDI 139

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASF 209
           :|   |||:||:|| |  |       |  |:| : |||  |::::|| :||:|||||  |
Db     140 TKHRPQVYMYAVPYQWIVPFDRIRDGVHAMVAQSVRRTPRSSIDPQVKNFQWGDLIRAVQ 199

Qy     210 EAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGI 269
           |  |||    :||| |  :||| |||| ::||| : || | ||||||||||  |:|: :|
Db     200 ETHDRGFEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTVLEIAESLGH 259

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
           || | |:|  || ||||::  ||||||  |  |:||  : :| |||:| :|  |:| |  |
Db     260 EAILADITLAELLDADEVLGCTTAGGVWPFVSVDGNPISDGVPGPITQSIIRRYWELNVE 319

Qy     330 PGPLIETIEY 339
             |:  ::|
Db     320 SSSLLTPVQY 329


RESULT 5
US-10-112-706-4
; Sequence 4, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
;  APPLICANT: YAMADA, Yukio
;  APPLICANT:  IWASAKI, Akira
;  APPLICANT:  KIZAKI, No. US20020192786Aliyuki
;  APPLICANT:  MATSUMOTO, Keiji
;  APPLICANT:  IKENAKA, Yasuhiro
;  APPLICANT:  OGURA, Masahiro
;  APPLICANT:  HASEGAWA, Junzo
;  TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
;  FILE REFERENCE: 1422-398P
;  CURRENT APPLICATION NUMBER: US/10/112,706
;  CURRENT FILING DATE:  2002-04-02
;  PRIOR APPLICATION NUMBER: 09/403,493
;  PRIOR FILING DATE: 1999-10-22
;  PRIOR APPLICATION NUMBER: JP 9-121732
;  PRIOR FILING DATE: 1997-04-23
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 330
;   TYPE: PRT
;   ORGANISM: Arthrobacter sp.
US-10-112-706-4

  Query Match             48.0%;  Score 847;  DB 9;  Length 330;
  Best Local Similarity   51.0%;  Pred. No. 2.7e-72;
  Matches  158;  Conservative   56;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      30 IQYSDYELDTSSPFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRL 89
           | ||||||| ::| ||| ||||| ::|  ||:||||| |: |||:|||| |||:|| |||
Db      21 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDQGYLHSDVTYTVFHVWNGNAFRL 80

Qy      90 GDHLDRLLDGASKLRLDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDL 149
             ||::||   |  :|:    ::||: ||  : |: ::||||:||::::|||    ||:|:
Db      81 DDHIERLFSNAESMRIIPPLTQDEVKEIALELVAKTELREAFVSVSITRGYSSTPGERDI 140

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASF 209
           :|   |||:||:|| |  |       |  |:| : |||  |::::|| :||:|||||  |
Db     141 TKHRPQVYMYAVPYQWIVPFDRIRDGVHAMVAQSVRRTPRSSIDPQVKNFQWGDLIRAVQ 200

Qy     210 EAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRNALPGITRKTVFELADQMGI 269
           |  |||    :||| |  :||| |||| ::||| : || | ||||||||||  |:|: :|
Db     201 ETHDRGFEAPLLLDGDGLLAEGSGFNVVVIKDGVVRSPGRAALPGITRKTVLEIAESLGH 260

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
           || | |:|  || ||||::  ||||||  |  |:||  : :| |||:| :|  |:| |  |
Db     261 EAILADITLAELLDADEVLGCTTAGGVWPFVSVDGNPISDGVPGPITQSIIRRYWELNVE 320

Qy     330 PGPLIETIEY 339
             |:  ::|
Db     321 SSSLLTPVQY 330


RESULT 6
US-10-027-450-29
; Sequence 29, Application US/10027450
; Patent No. US20020102715A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT:  Hitz, William D.
;  APPLICANT:  Kinney, Anthony J.
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/10/027,450
;  CURRENT FILING DATE:  2001-12-20
;  PRIOR APPLICATION NUMBER: 60/063,423
;  PRIOR FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 29
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Methanococcus jannaschii
US-10-027-450-29

  Query Match             21.8%;  Score 385;  DB 12;  Length 288;
  Best Local Similarity   31.9%;  Pred. No. 1.5e-28;
  Matches   90;  Conservative   58;  Mismatches  116;  Indels   18;  Gaps    5;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :: |:::  ::||:|:|| |  : |  :     : | :| | :|:||| | |  | :|
Db       4 YLNGKFVDEKDAKVSVFDHGLLYGDGVFEGIRAYDGVVFMLKEHIDRLYDSAKSLCIDIP 63

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYG-----KRKGEKDLSKLTHQVYIYAIPY 163
             :|:|: ::  : : :: ||:::: | |||| |      ||  |        ::  |||
Db      64 LTKEEMIDVVLETLRVNNLRDAYIRLVVTRGVGDLGLDPRKCGKP------TIFCIAIP- 116

Qy     164 LWAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAILLD 223
             ||      |  ||    |||   : ::| :|:  : :   |  :|   |    | |||
Db     117 ---MPPLLGEDGIRAITV-SVRRLPVDVLNPAVKSLNYLNSVLAKIQANYAGVDEAFLLD 172

Qy     224 SDNCVAEGPGFNVCIVKDGKLASPS--RNALPGITRKTVFELADQMGIEATLRDVTSREL 281
           | || | |: |||:| | :|   :: | ||||  | :|| : |||     :|  :|
Db     173 DKGFVVEGTGDNIFIVKNGVLKTPPVYQSILKGITRDVVIKLAKEEGIEVVEEPLTLHDL 232

Qy     282 YDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           | ||||    ||  : |:  :||  : | : | :|  ::::|
Db     233 YTADELFITGTAAEIVPVFEIDGRVINNKQVGEITKKLKEKF 274
```

```
RESULT 7
US-10-027-450-31
; Sequence 31, Application US/10027450
; Patent No. US20020102715A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT:  Hitz, William D.
;  APPLICANT:  Kinney, Anthony J.
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/10/027,450
;  CURRENT FILING DATE:  2001-12-20
;  PRIOR APPLICATION NUMBER: 60/063,423
;  PRIOR FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 31
;   LENGTH: 310
;   TYPE: PRT
;   ORGANISM: Zea mays
US-10-027-450-31

  Query Match             18.3%;  Score 322.5;  DB 12;  Length 310;
  Best Local Similarity   29.0%;  Pred. No. 1.5e-22;
  Matches   82;  Conservative   54;  Mismatches  134;  Indels   13;  Gaps    5;

Qy      47 VAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           : |:  | ||   ||:|:||:     |  :    :: | :|:| :||||| | |  :
Db      18 LVWVGDELLPRNSAKVSVFDSVVQGGDAVWEGLRIYDGKVFKLDEHLDRLFDSAKAMAFS 77

Qy     107 AGYSKDELAE-ITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL- 164
             ::| : : | |  ::       : : ||:||       |:|  | ::    :|    :
Db      78 NVPTRDWIKDAIFKTLIANGMFNNAHIRLTLTR------GKKVTSGMSPAFNLYGCALIV 131

Qy     165 ---WAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWGDLTAASFEAKDRGARTAIL 221
               |  |  :   |  :|    ||   |::|| | :    :   |  |     |  ||:
Db     132 LAEWKPPVYDNSHG-IKLVTATTRRNSPNSIDPKIHHNNLINNILAKIEGNLAQAEDAIM 190

Qy     222 LDSDNCVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVFELADQMGIEATLRDVTSRE 280
           || |  |:|    |: :|| | : :|  :  ||||||| || :|  :        | ::  |
Db     191 LDKDGFVSETNATNIFMVKKGIVLTPHADYCLPGITRATVMDLVVKENFVLHERRISLSE 250

Qy     281 LYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           : |||:    | | :||:  :||  :|:|: ||:|  |:  :
Db     251 FHAADEVWTTGTMGEITPVVMIDGREIGDGKIGPVTRQIQKAY 293


RESULT 8
US-10-060-432-32
; Sequence 32, Application US/10060432
; Publication No. US20030040092A1
;   GENERAL INFORMATION:
;        APPLICANT: WARREN, Patrick V.
;                   SWANSON, Ronald V.
;        TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
;        NUMBER OF SEQUENCES: 32
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
;                        CECCHI, STEWART & OLSTEIN
;             STREET: 6 BECKER FARM ROAD
;             CITY: ROSELAND
;             STATE: NEW JERSEY
;             COUNTRY: USA
;             ZIP: 07068
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 INCH DISKETTE
;             COMPUTER: IBM PS/2
;             OPERATING SYSTEM: MS-DOS
;             SOFTWARE: WORD PERFECT 5.1
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/060,432
;             FILING DATE: 29-Jan-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/481,733
;             FILING DATE: 11-JAN-2000
;             APPLICATION NUMBER: US/08/599,171
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: HERRON, CHARLES J.
;             REGISTRATION NUMBER: 28,019
;             REFERENCE/DOCKET NUMBER: 331400-38
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 201-994-1700
;             TELEFAX: 201-994-1744
;   INFORMATION FOR SEQ ID NO: 32:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 303 AMINO ACIDS
;             TYPE: AMINO ACID
;             TOPOLOGY: LINEAR
;        MOLECULE TYPE: PROTEIN
;        SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-10-060-432-32

  Query Match             17.9%;  Score 315;  DB 9;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 7.3e-22;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :       :     :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIFEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||       :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
             |   : :  |           ||   : |  ::      |   ||: | :     :
Db     113 RNLEVSLAVIVFP----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | ||||  |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| ||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
           : :
Db     281 AELY 284


RESULT 9
US-09-905-173-32
; Sequence 32, Application US/09905173
; Patent No. US20020132295A1
; GENERAL INFORMATION:
;  APPLICANT: DIVERSA CORPORATION
;  APPLICANT:  SHORT, Jay M.
;  APPLICANT:  WARREN, Patrick V.
;  APPLICANT:  SWANSON, Ronald V.
;  APPLICANT:  MATHUR, Eric J.
;  TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND
;  TITLE OF INVENTION:  USE THEREOF
;  FILE REFERENCE: DIVER1240-7
;  CURRENT APPLICATION NUMBER: US/09/905,173
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: US 09/412,184
;  PRIOR FILING DATE: 1999-10-04
;  PRIOR APPLICATION NUMBER: US 09/389,537
;  PRIOR FILING DATE: 1999-09-02
;  PRIOR APPLICATION NUMBER: US 08/646,590
```

```
;  PRIOR FILING DATE: 1996-05-08
;  PRIOR APPLICATION NUMBER: US 08/599,171
;  PRIOR FILING DATE: 1996-02-09
;  PRIOR APPLICATION NUMBER: US 09/481,733
;  PRIOR FILING DATE: 2000-01-11
;  PRIOR APPLICATION NUMBER: US 09/069,226
;  PRIOR FILING DATE: 1998-04-27
;  NUMBER OF SEQ ID NOS: 40
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: Pyrobaculum aerophilum
US-09-905-173-32

  Query Match             17.9%;  Score 315;  DB 10;  Length 303;
  Best Local Similarity   28.0%;  Pred. No. 7.3e-22;
  Matches   85;  Conservative   62;  Mismatches  113;  Indels   44;  Gaps   11;

Qy      42 PFAGGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYT-VAHVWHGN---IFRLGDHLDRLL 97
           |:|  : |::|  |  |:||| :      :    :  :   |:|:   :||| :|:||:
Db       3 PYAKYI-WLDGRILKWEDAKIHVLTHALHYGTSIFEGIRGYWNGDNLLVFRLEEHIDRMY 61

Qy      98 DGASKLRLDAGYSKDELAEITKKCVSMSQLRE-------SFV-NLTVTRGYGKRKGEKDL 149
             |  | ::  |:::|: :   : :  :  ||       :|| : |||          |:
Db      62 RSAKILGINIPYTREEVRQAVLETIKANNFREDVYIRPVAFVASQTVT---------LDI 112

Qy     150 SKLTHQVYIYAIPYLWAFPPAEQIFGTTAIVPRHVR------RAGRNTVDPTIKNYQ--W 201
           |   : :  |              ||   : |  ::      |   ||: | :      :
Db     113 RNLEVSLAVIVFP-----------FG-KYLSPNGIKATIVSWRRVHNTMLPVMAKIGGIY 160

Qy     202 GDLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASP--SRNALPGITRKT 259
           :   |  ||: ||   |:|:| :  | || | |: ||: |:| :|    : | |||| |
Db     161 VNSVLALVEARSRGFDEALLMDVNGYVVEGSGENIFIVRGGRLFTPPVHESILEGITRDT 220

Qy     260 VFELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
           | :|:  :|:    : :|  |:| |||:  | ||  :||:  :||  :| |:|||:|  |
Db     221 VIKLSGDVGLRVEEKPITREEVYTADEVFLVGTAAEITPVVEVDGRTIGTGKPGPITTKI 280

Qy     320 RDRF 323
           : :
Db     281 AELY 284


RESULT 10
US-10-136-253-6
; Sequence 6, Application US/10136253
; Patent No. US20020136737A1
; GENERAL INFORMATION:
;  APPLICANT: FRANKEL, Fred R
;  APPLICANT:  PORTNOY, Daniel A.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS COMPRISING LISTERIA DAL AND DAT GENES
;  FILE REFERENCE: 053893-5011-02
;  CURRENT APPLICATION NUMBER: US/10/136,253
;  CURRENT FILING DATE:  2002-05-01
;  PRIOR APPLICATION NUMBER: 09/520,207
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: PCT/US98/24357
;  PRIOR FILING DATE: 1998-11-13
;  PRIOR APPLICATION NUMBER: 08/972,902
;  PRIOR FILING DATE: 1997-11-18
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 289
;   TYPE: PRT
;   ORGANISM: Listeria monocytogenes
US-10-136-253-6

  Query Match             16.4%;  Score 289;  DB 12;  Length 289;
  Best Local Similarity   28.9%;  Pred. No. 2e-19;
  Matches   83;  Conservative   51;  Mismatches  123;  Indels   30;  Gaps    6;

Qy      50 IEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGY 109
           :   :  |:| : | | |:   | | |  :::|  |   :|:|||   |:|: |   |
Db       5 VNNHLVEREDATVDIEDRGYQFGDGVYEVVRLYNGKFFTYNEHIDRLYASAAKIDLVIPY 64

Qy     110 SKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPP 169
           ||:|| |: :| |: : :    | | ||||    :               : ||    ||
Db      65 SKEELRELLEKLVAENNINTGNVYLQVTRGVQNPRN------------HVIPD--DFPL 109

Qy     170 AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGART 218
              :      |||: |:        | :: :|           |:: | : :|  : |
Db     110 EGVLTAAAREVPRNERQFVEGGTAITEEDVRWLRCDIKSLNLLGNILAKN-KAHQQNALE 168

Qy     219 AILLDSDNCVAEGPGFNVCIVKDGKLASPSRN--ALPGITRKTVFELADQMGIEATLRDV 276
           |||   :  | |    || |:||| | : : :   | ||||: : ::| : ||     |
Db     169 AILHRGEQ-VTECSASNVSIIKDGVLWTHAADNLILNGITRQVIIDVAKKNGIPVKEADF 227

Qy     277 TSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |  :| :|||:   :|   :|||  :||  | :|: ||:|  :   |
Db     228 TLTDLREADEVFISSTTIEITPITHIDGVQVADGKRGPITAQLHQYF 274


RESULT 11
US-10-136-253-7
; Sequence 7, Application US/10136253
; Patent No. US20020136737A1
; GENERAL INFORMATION:
;  APPLICANT: FRANKEL, Fred R
;  APPLICANT:  PORTNOY, Daniel A.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS COMPRISING LISTERIA DAL AND DAT GENES
;  FILE REFERENCE: 053893-5011-02
;  CURRENT APPLICATION NUMBER: US/10/136,253
;  CURRENT FILING DATE:  2002-05-01
;  PRIOR APPLICATION NUMBER: 09/520,207
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: PCT/US98/24357
;  PRIOR FILING DATE: 1998-11-13
;  PRIOR APPLICATION NUMBER: 08/972,902
;  PRIOR FILING DATE: 1997-11-18
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 282
;   TYPE: PRT
;   ORGANISM: Listeria monocytogenes
US-10-136-253-7

  Query Match             16.0%;  Score 283;  DB 12;  Length 282;
  Best Local Similarity   25.3%;  Pred. No. 7.1e-19;
  Matches   73;  Conservative   59;  Mismatches  124;  Indels   32;  Gaps    6;

Qy      49 WIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAG 108
           :| ||::   |||:|  | |:   |  |     : | :| : :| :|  :  ||:::|| |
Db       5 FINGEFIDQNEAKVSYEDRGYVFGDGIYEYIRAYDGKLFTVTEHFERFIRSASEIQLDLG 64

Qy     109 YSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFP 168
           |: :|| :: :: : :: ::   : :  |||   |        ::   :  :|  |   :
Db      65 YTVEELIDVVRELLKVNNIQNGGIYIQATRGVAPRNHSFPTPEVKPVIMAFAKSYDRPYD 124

Qy     169 PAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW----------GDLTAASFEAKDRGAR 217
           |                  | |    |::: :|           |:: |  :  |
Db     125 DLEN--------------GINAA--TVEDIRWLRCDIKSLNLLGNVLAKEYAVKYNAGE 167

Qy     218 TAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEATLRD 275
           :      | ||   ||  :||| :   |  |  | ||||| :  :::   |
Db     168 A--IQHRGETVTEGASSNVYAIKDGAIYTHPVNNYILNGITRKVIKWISEDEDIPFKEET 225

Qy     276 VTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRF 323
           |   | :|||::  :|:  |||:  :||| ||:|: ||:|  ::: |
Db     226 FTVEFLKNADEVIVSSTSAEVTPVVKIDGEQVGDGKVGPVTRQLQEGF 273


RESULT 12
US-10-136-253-9
```

```
; Sequence 9, Application US/10136253
; Patent No. US20020136737A1
; GENERAL INFORMATION:
;  APPLICANT: FRANKEL, Fred R
;  APPLICANT:  PORTNOY, Daniel A.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS COMPRISING LISTERIA DAL AND DAT GENES
;  FILE REFERENCE: 053893-5011-02
;  CURRENT APPLICATION NUMBER: US/10/136,253
;  CURRENT FILING DATE:  2002-05-01
;  PRIOR APPLICATION NUMBER: 09/520,207
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: PCT/US98/24357
;  PRIOR FILING DATE: 1998-11-13
;  PRIOR APPLICATION NUMBER: 08/972,902
;  PRIOR FILING DATE: 1997-11-18
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
;   LENGTH: 283
;   TYPE: PRT
;   ORGANISM: Listeria monocytogenes
US-10-136-253-9

  Query Match             15.8%;  Score 278.5;  DB 12;  Length 283;
  Best Local Similarity   27.6%;  Pred. No. 1.9e-18;
  Matches   83;  Conservative   50;  Mismatches  131;  Indels   37;  Gaps    9;

Qy      45 GGVAWIEGEYLPAEEAKISIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLR 104
           |   |   : :  || ||   | |:   |  | |  |::| :| : :|:|||   | |:|
Db       2 GYTLW-NDQIVKDEEVKIDKEDRGYQFGDGVYEVVKVYNGEMFTVNEHIDRLYASAEKIR 60

Qy     105 LDAGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYL 164
           :   |:||:  ::  : |  ::|    :   ||||   |          ||
Db      61 ITIPYTKDKFHQLLHELVEKNELNTGHIYFQVTRGTSPR---------AHQ--------- 102

Qy     165 WAFPP---AEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQW--GDLTA--------ASFEA 211
             ||       | | |   || :    :      ::: :|   |: :        |  ||
Db     103 --FPENTVKPVIIGYTKENPRPLENLEKGVKATFVEDIRWLRCDIKSLNLLGAVLAKQEA 160

Qy     212 KDRGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGI 269
           ::|   || |  :| | ||   ||  :||| |   |: |  | |||| |   |::: :
Db     161 HEKGCYEAI-LHRNNTVTEGSSSNVFGIKDGILYTHPANNMILKGITRDVVIACANEINM 219

Qy     270 EATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWALMDE 329
              |: |    |||   :|   :||:  :||: : :|: |  |  :: :|   : :
Db     220 PVKEIPFTTHEALKMDELFVTSTTSEITPVIEIDGKLIRDGKVGEWTRKLQKQFETKIPK 279

Qy     330 P 330
           |
Db     280 P 280


RESULT 13
US-10-136-253-8
; Sequence 8, Application US/10136253
; Patent No. US20020136737A1
; GENERAL INFORMATION:
;  APPLICANT: FRANKEL, Fred R
;  APPLICANT:  PORTNOY, Daniel A.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS COMPRISING LISTERIA DAL AND DAT GENES
;  FILE REFERENCE: 053893-5011-02
;  CURRENT APPLICATION NUMBER: US/10/136,253
;  CURRENT FILING DATE:  2002-05-01
;  PRIOR APPLICATION NUMBER: 09/520,207
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: PCT/US98/24357
;  PRIOR FILING DATE: 1998-11-13
;  PRIOR APPLICATION NUMBER: 08/972,902
;  PRIOR FILING DATE: 1997-11-18
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 283
;   TYPE: PRT
;   ORGANISM: Listeria monocytogenes
US-10-136-253-8

  Query Match             14.2%;  Score 250;  DB 12;  Length 283;
  Best Local Similarity   26.9%;  Pred. No. 9.7e-16;
  Matches   79;  Conservative   49;  Mismatches  110;  Indels   56;  Gaps    9;

Qy      58 EEAKISIF--DTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLDAGYSKDELA 115
           ||  |:|   | |:   |  | |  |::|::|   :|:||    | |:||   |:|| |
Db      12 EEGSITISPEDRGYQFGDGIYEVIKVYNGHMFTAQEHIDRFYASAEKIRLVIPYTKDVLH 71

Qy     116 EITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLSKLTHQVYIYAIPYLWAFPPAEQIFG 175
           ::    :  : |    |   :|||   |                        || |
Db      72 KLLHDLIEKNNLNTGHVYFQITRGTTSRN-------------------HIFPDA----S 107

Qy     176 TTAIVPRHVRRAGRNTVDPTIKNYQWG-------------DLTA--------ASFEAKD 213
             |::  :|:     | : :|:|:: |               |: :        |  || :
Db     108 VPAVLTGNVK-----TGERSIENFEKGVKATLVEDVRWLRCDIKSLNLLGAVLAKQEASE 162

Qy     214 RGARTAILLDSDNCVAEGPGFNVCIVKDGKL-ASPSRN-ALPGITRKTVFELADQMGIEA 271
           :|   |||   | : |    ||  :|||||   |: |  | ||||: : : | :: :
Db     163 KGCYEAILHRGD-IITECSSANVYGIKDGKLYTHPANNYILNGITRQVILKCAAEINLPV 221

Qy     272 TLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRDRFWA 325
             :|  :|   ||::  : :  |||:  :||: :| | ||  |  ::  | |
Db     222 IEEPMTKGDLLTMDEIIVSSVSSEVTPVIDVDGQQIGAGVPGEWTRKLQKAFEA 275


RESULT 14
US-10-027-450-24
; Sequence 24, Application US/10027450
; Patent No. US20020102715A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT:  Hitz, William D.
;  APPLICANT:  Kinney, Anthony J.
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/10/027,450
;  CURRENT FILING DATE:  2001-12-20
;  PRIOR APPLICATION NUMBER: 60/063,423
;  PRIOR FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 24
;   LENGTH: 285
;   TYPE: PRT
;   ORGANISM: Glycine max
US-10-027-450-24

  Query Match             13.5%;  Score 238;  DB 12;  Length 285;
  Best Local Similarity   28.2%;  Pred. No. 1.3e-14;
  Matches   81;  Conservative   48;  Mismatches  116;  Indels   42;  Gaps   10;

Qy      51 EGEYLPAEEAKI----SIFDTGFGHSDLTYTVAHVWHGNIFRLGDHLDRLLDGASKLRLD 106
           |   :| ::  :      :|||           | :  | :: |  |||| |  ||  ::|
Db       2 EAMVIPMDDHMVHRGHGVFDT----------AAIMDGYLYELDQHLDRFLRSASMSKID 50

Qy     107 AGYSKDELAEITKKCVSMSQLRESFVNLTVTRGYGKRKGEKDLS-KLTHQVYIYAIPYLW 165
             : :  :  |  : || |: |:  :   :: |      |:  ||    |:  :|||
Db      51 PPFDRGSIRRILIQTVSASKCRKGSLRYWLSAG----PGDFQLSPSCCHRSSLYAIVIQD 106

Qy     166 AFPPAEQIFG----TTAIVPRHVRRAGRNTVD--PTIKNYQWGDLTAASFEAKDRGARTA 219
           | :    |    |::|  :| : |   :|:  | :                :  ||:: ||
Db     107 LSPSSPNFRGVKVVTSSIPIKHPKFAITKSVNYLPNV---------LSKVEAEEAGAFVG 157

Qy     220 ILLDSDNCVAEGPGFNVCIV-KDGKLASPS-RNALPGITRKTVFELADQMGIEATL---- 273
           | || :  |||||  ||  | || :|  |     | | | | |  ||: :  |  |
Db     158 IWLDGEGFVAEGPNMNVAFVTKDKELIMPHFDKILSGCTAKRVLTLAESLLREGKLKGIR 217
```

```
Qy    274 -RDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAI 319
          : ||  |   |||:| : :   | |:   | : :|:|: ||:| |:
Db    218 VKTVTVEEGKQADEMMLLGSGVLVCPVVQWDEQVIGDGKEGPITQAL 264


RESULT 15
US-10-027-450-33
; Sequence 33, Application US/10027450
; Patent No. US20020102715A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT:  Hitz, William D.
;  APPLICANT:  Kinney, Anthony J.
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/10/027,450
;  CURRENT FILING DATE:  2001-12-20
;  PRIOR APPLICATION NUMBER: 60/063,423
;  PRIOR FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 33
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Triticum aestivum
US-10-027-450-33

  Query Match            12.8%;  Score 226.5;  DB 12;  Length 255;
  Best Local Similarity   28.6%;  Pred. No. 1.4e-13;
  Matches   70;  Conservative   43;  Mismatches  119;  Indels   13;  Gaps    5;

Qy     88 RLGDHLDRLLDGASKLRLDAGYSKDELAE-ITKKCVSMSQLRESFVNLTVTRGYGKRKGE 146
          :| :||||| |    :      |:| : : | |   :       : : ||:||      |:
Db      4 KLEEHLDRLFDSTKAMAFSNVPSRDWIKDAIFKTLNANGMFNNAHIRLTLTR------GK 57

Qy    147 KDLSKLTHQVYIYAIPYL----WAFPPAEQIFGTTAIVPRHVRRAGRNTVDPTIKNYQWG 202
          |  | ::    :|    :    |  |  :   |   :|    ||   |:||  | :
Db     58 KVTSGMSPTFNLYGCVLIVLAEWKPPVYDNSHG-IKLVTAATRRNSPNSVDSKIHHNNLI 116

Qy    203 DLTAASFEAKDRGARTAILLDSDNCVAEGPGFNVCIVKDGKLASPSRN-ALPGITRKTVF 261
          :   |  |     |  ||:|| |  |:|    |: :|| |.: :|  :  ||||| ||
Db    117 NNILAKIEGNLAQAEDAIMLDQDGFVSETNATNIFMVKKGIVLTPHADYCLPGITRATVK 176

Qy    262 ELADQMGIEATLRDVTSRELYDADELMAVTTAGGVTPINSLDGEAVGNGEPGPLTVAIRD 321
          :|  :  :    | ::  | : |||:    | | :||:  :||  :|:|: | :|  |:
Db    177 DLVVKENLVLHERRISLSEFHAADEVWTTGTMGEITPVVMIDGREIGDGKIGLVTRQIQS 236

Qy    322 RFWAL 326
          :  |
Db    237 AYKVL 241


Search completed: April 17, 2003, 17:44:44
Job time : 21 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 20:06:42 ; Search time 2799 Seconds
(without alignments)
10605.524 Million cell updates/sec

Title: US-10-067-291-2
Perfect score: 1020
Sequence: 1 atgactgctctttcagacct.....tcgaaacgatcgaatactga 1020

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1020	100.0	1020	6	BD000505	BD000505 Protein,
	2	247	24.2	329709	1	AP002997	AP002997 Mesorhizo
c	3	191.6	18.8	307551	1	AP003015	AP003015 Mesorhizo
c	4	153.2	15.0	149050	1	MLO672114	AL672114 Mesorhizo
c	5	129.4	12.7	10364	1	AE005957	AE005957 Caulobact
c	6	126.6	12.4	10685	1	AE010451	AE010451 Methanopy
	7	109.8	10.8	190050	1	AL646080	AL646080 Ralstonia
c	8	95.8	9.4	12730	1	AE005101	AE005101 Halobacte
c	9	95	9.3	348077	1	AP003000	AP003000 Mesorhizo
c	10	87.2	8.5	212050	1	AL646060	AL646060 Ralstonia
c	11	85.8	8.4	17576	1	AE004913	AE004913 Pseudomon
	12	84	8.2	10993	1	AE011153	AE011153 Methanosa
	13	83.2	8.2	7991	1	PAU63816	U63816 Pseudomonas
	14	82.6	8.1	10196	1	AE004996	AE004996 Halobacte
	15	72.4	7.1	10029	1	AE013330	AE013330 Methanosa
	16	72.4	7.1	23533	1	AE001039	AE001039 Archaeogl
	17	70	6.9	2748	1	SAUSIGA	M94370 Stigmatella
c	18	68.8	6.7	12026	1	AE005109	AE005109 Halobacte
	19	68.6	6.7	31226	1	SC6F7	AL353870 Streptomy
	20	67.8	6.6	14991	1	AE005125	AE005125 Halobacte
	21	67.2	6.6	347660	1	AP002994	AP002994 Mesorhizo
	22	64.6	6.3	11202	1	AE005123	AE005123 Halobacte
	23	64.2	6.3	12510	1	AE005848	AE005848 Caulobact
	24	64.2	6.3	77457	1	AF210249	AF210249 Streptomy
	25	63.2	6.2	30000	6	AX250261	AX250261 Sequence
	26	62.8	6.2	1173	1	AF246405	AF246405 Acidithio
	27	62.8	6.2	11170	1	AE005047	AE005047 Halobacte
c	28	62.8	6.2	40544	1	SCH5	AL035636 Streptomy
	29	62.2	6.1	10112	1	AE012915	AE012915 Chlorobiu
c	30	62.2	6.1	23404	1	SC10A9	AL583943 Streptomy
	31	62	6.1	20107	1	AE008882	AE008882 Salmonell
	32	62	6.1	96086	1	STYSTMD1	AF233324 Salmonell
	33	61.2	6.0	45313	1	SCD95A	AL357432 Streptomy
	34	60.8	6.0	2475	15	AF039028	Af039028 Streptomy
c	35	60.8	6.0	15901	1	SC3A3	AL109849 Streptomy
	36	60.8	6.0	32039	1	SC7A1	AL034447 Streptomy
	37	60.8	6.0	69301	1	STU82965	U82965 Streptomcye
c	38	60.4	5.9	10652	1	AB006206	AB006206 Streptomy
	39	60.4	5.9	34593	1	SC2K8	AL589164 Streptomy
	40	60.2	5.9	6825	1	FSU53363	U53363 Frankia sp.
	41	60	5.9	23030	1	SC9E12	AL391751 Streptomy
	42	59.6	5.8	3522	1	AF421216	AF421216 Streptomy
c	43	59.4	5.8	31624	1	SCD63	AL161755 Streptomy
c	44	59.2	5.8	54101	1	AB070956	AB070956 Streptomy
	45	58.4	5.7	9169	6	AX427931	AX427931 Sequence

ALIGNMENTS

RESULT 1
BD000505
LOCUS BD000505 1020 bp DNA linear PAT 31-JAN-2002
DEFINITION Protein, its gene and use thereof.
ACCESSION BD000505
VERSION BD000505.1 GI:18623618
KEYWORDS JP 2000342276-A/1.
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1020)
AUTHORS Takashima,Y., Wieser,M. and Koda,K.
TITLE Protein, its gene and use thereof
JOURNAL Patent: JP 2000342276-A 1 12-DEC-2000;
SUMITOMO CHEMICAL CO LTD

```
COMMENT     OS   Mycobacterium aurum SC-S423
            PN   JP 2000342276-A/1
            PD   12-DEC-2000
            PF   17-MAR-2000 JP 2000075752
            PR
            PI   YOSHIKI TAKASHIMA,MARKOS WIESER,KEN KODA
            PC   C12N15/09,C07K14/35,C12N1/15,C12N1/19,C12N1/21,C12N5/10, PC
            C12P13/00,
            PC   C12P21/02//C07M7:00,C12N15/00,C12N5/00
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(1020).
FEATURES            Location/Qualifiers
     source         1. .1020
                    /organism="unidentified"
                    /db_xref="taxon:32644"
BASE COUNT      198 a    350 c    313 g    159 t
ORIGIN

  Query Match           100.0%;  Score 1020;  DB 6;  Length 1020;
  Best Local Similarity  100.0%;  Pred. No. 7e-131;
  Matches 1020;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60

Qy      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120

Qy     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180

Qy     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240

Qy     241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300

Qy     301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360

Qy     361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420

Qy     421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480

Qy     481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540

Qy     541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600

Qy     601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660

Qy     661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720

Qy     721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780

Qy     781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840

Qy     841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900

Qy     901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960

Qy     961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020


RESULT 2
AP002997
LOCUS       AP002997               329709 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 4/21.
ACCESSION   AP002997 BA000012
VERSION     AP002997.2  GI:14022051
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000).
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 329709)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994965.
FEATURES            Location/Qualifiers
     source         1. .329709
                    /organism="Mesorhizobium loti"
                    /strain="MAFF303099"
                    /db_xref="taxon:381"
     gene           complement(79. .402)
                    /gene="mll1244"
     CDS            complement(79. .402)
                    /gene="mll1244"
                    /note="unknown protein"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="BAB48662.1"
                    /db_xref="GI:14022052"
                    /translation="MSRKGLSIPFNGSFHYWADCLMVQRPIMSDLLLSSIFTAFTMVR
                    LLKGPWLRNPQYLATGILGAVVAMLLLNGLWPAYDDDFVIGGVTGIFGSWAGMALFDA
                    ILGIA"
     gene           411. .2480
                    /gene="mlr1245"
     CDS            411. .2480
                    /gene="mlr1245"
                    /codon_start=1
                    /transl_table=11
```

/product="probable electron transfer protein"
/protein_id="BAB48663.1"
/db_xref="GI:14022053"
/translation="MNSPANITDPLVLFMPSGKRGRFPVGTPVLDAARQLGVYVESVC
GGRATCGRCQIEVQEGNFAKHKIVSSNDHISPKGPKEERYERVRGLPERRRLSCSAQI
LGDLVIDVPQDTVINAQTIRKDADTRVIARDTAIRMCYVEIEEPDMHKPLGDLDRLKI
ALMKDWGFKNLEFDFYLLPQVQGILRKGNWTATAAIHKDADSDIARVIALWPGLKNEA
YGLACDIGSTTIAMHLVSLLSGRVAASSGTSNPQIRFGEDLMSRVSYVMMNPDGREGM
TVAVREAISSLVDKVCAEGNVQRNDILDSVFVGNPIMHHLFLGIDPTELGGAPFALAV
SGAVRIKASDIGLKLNQGARLYMLPCIAGHVGADAAAVTLSEGPHRQDEMMLIVDVGT
NAEIVLGNRARVVAASSPTGPAFEGAEISGGQRAAPGAIERVRIDPDTLEPKYRVIGS
ELWSDEPGFLDSVQATGVTGICGSGIIEIVAEMYLAGIISEDGVVDGSLAARSPRVTA
NGRTFSYVLKEGEPKITITQTDVRAIQLAKAALYAGTKLLMEKQNTEHVDRIHFAGAF
GSFIDPKYAMVLGLIPDCDLDKVSAVGNAAGAGARMALLNRGYRREIEETVSQIEKIE
TALEPKFQEHFVYAMALPNKVDPFPKLSAAVKLPPRKTVSEDGIAGDATPRRRSREGH
AARRSRE"
gene 2529. .2852
/gene="mlr1246"
CDS 2529. .2852
/gene="mlr1246"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48664.1"
/db_xref="GI:14022054"
/translation="MFPDRTDPETETLFSRNLSDTWQAHLPRIERRPEIILGRRSSRE
MKMSMFENLGRFGVTIRQNHARNKAIRALNSLPAHVQKDIGWPASPPSDPQAALASLL
LGTAR"
gene 3087. .4040
/gene="mlr1247"
CDS 3087. .4040
/gene="mlr1247"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="BAB48665.1"
/db_xref="GI:14022055"
/translation="MPSLKSHFVSFVLRYSRKKAFSSPENLRRWIAAARKTEDHHPPA
SLHQRLHIQTRTVDGFPVYEIAPKAGEQRRILYLHGGAYVFEITPYHWHLIADMADRL
GYGITVPIYPIAPEHDFHAMFGMVGDVYRQMLDETEAEDIVFMGDSAGGNMAVVLTMM
AAEEGLPLPARHVLISPGLDMSLSNPKLFEAERNDPWLGIAGGLEAIRMYSAGIDRSD
WHISPIYGDVSVLPTTLLLTGSHDLLTPDNLIFAQKARDAGVEVEVVHEEGMFHAWPL
IDMPEARRARQHIVAFLGENRSPVRQDGKTRFQAPSAEAAE"
gene complement(4062. .4961)
/gene="mll1248"
CDS complement(4062. .4961)
/gene="mll1248"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="BAB48666.1"
/db_xref="GI:14022056"
/translation="MNAPLNHPLPLLDLDVLRTFVAIAETGSFTTAANAVFRTPSAVS
MQIKKLEDILGRSVFARDARSVSLTTDGEMLLGYARRLLSINREVVSKFIIPEIVGVV
RLGSPDDYGERVLPHVLKRFAQSHPSIAVDVTIDQSSNLRRRMDDRALDITLLTNSYK
TSALGAEVLLTEPIVWAGAKGGCAHLREPLPVSLWEEGCAWRAGALEALGREGRNYRV
AYMSAHTAGQRAAIMADLAVAPLPRSFLGNDMVELCPKDGMPDIGTYNLAMVVAPDAS
APVKAVADHIRATFEVFRETGKF"
gene 5113. .5379
/gene="msr1249"
CDS 5113. .5379
/gene="msr1249"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48667.1"
/db_xref="GI:14022057"
/translation="MGLKHAACKTSQIKRRNVPGPAPKENEMSILSSLGRIATEFSAA
RARYQTERAIRSLPIELQKDIGWPEASDTKTGTRNGVGSWAGAK"
gene 6136. .7137
/gene="mlr1251"
CDS 6136. .7137
/gene="mlr1251"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="BAB48668.1"
/db_xref="GI:14022058"
/translation="MNAIKHPIKRSFVFFLVPDFTMIAFATALDPLRSANRMLGYEAY
RWRLASIDGKPVRASNGVECAVNTSLEEERKKMAGPDRPNMAIVCSGVNVERYQNKSA
FAWLREEYNRGVAVGGLCTGAHILAAAGLLSNKRCAIHWENLPGFSEAFPKANVFADL
FEVDQNIYTCAGGTAALDMMLKLIGDDFDESLVNRVCEQVLTDRVRSPTDRQRLPLRA
RLGVQNSKVLTIIELMEGNLSEPLSLIEIADHVDLSRRQIERLFRTEMGRSPARYYLE
IRLDRARHLLIQSSMPVVEVAVACGFVSASHFSKCYRELYARSPQQERVDRKQLLAA"
gene complement(7153. .7956)
/gene="mll1253"
CDS complement(7153. .7956)
/gene="mll1253"
/codon_start=1
/transl_table=11
/product="ATP-binding protein of ABC transporter"
/protein_id="BAB48669.1"
/db_xref="GI:14022059"
/translation="MEPGVSKLLIEGVSRTFAGVRGGLPVKALMPIDLAVAANDFITI
LGPSGCGKSTLLRIVAGLEAPSEGRVLLDDKAVTRPGPDRGMVFQSYTLFPWLTVSEN
IAFGLRERGMAPKERDDIVASYVDLVGLKGFENHWPKQLSGGMQQRTAIARALANDPE
ILLLDEPFGALDNQTRGLMQELLLGIWERRKKTVLFVTHDIEEAIFMASRVIVMTARP
GSIKSDVAIDLPHPRHYTLKTSPEFSALKARLTEDIRVEAMRTAQAQPA"
gene complement(7941. .8720)
/gene="mll1254"
CDS complement(7941. .8720)
/gene="mll1254"
/codon_start=1
/transl_table=11
/product="permease protein of ABC transporter"
/protein_id="BAB48670.1"
/db_xref="GI:14022060"
/translation="MKMRPLHPVSPGLRTGLGISFFVLFIAFWAWITLGGHVNRIFLA
DPLSMLRDGWRLLVEDRFWLDILITIWRVFGGFVLASVVAVPIGIAMGAWKPVEAFLE
PFVSFARYLPASAFIPLLILWAGIGELEKLLVIFVGSVFQIILIVAVKVGSTRRDLVE
AAYTLGSTNNGIVKRVIMPANAPEIAETLRLVLGWAWTYVIVAELIGSSSGIGYMIIN
SQSRLATGQIIFGIIVIGLIGLLSDFAFKAFNRWLFPWSLA"
gene complement(8802. .9752)
/gene="mll1256"
CDS complement(8802. .9752)
/gene="mll1256"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48671.1"
/db_xref="GI:14022061"
/translation="MNRIAVFAATLTFAAGLSAPVMAATSVTIGISGWTGFAPLTLAK
QAGLFEKHGLDVTLKKVPQASRPLAIASGDLQCAATTVETWLVWNASGVTTKQIFQLD
KSYGADGIVVRNDIKTVADLKGKNVASSAPGTSPYFLLAWVLNKNGMSTKDVTVVNLE
PDAAAQAFLAGQNDAAVTYEPFISAVRDKADQGHILATTLDYPMVLDTVGCTPEFLKA
NPDAAKALADSYFEALDLIKKDPQKSYEIMGADVKQSAKEFEDSAKYLKWADKADNQQ
FFTKEFQDFSKTAGALLLQMGLIKEAPDVATLADTSAVAN"
gene complement(9963...11000)
/gene="mll1257"
CDS complement(9963. .11000)
/gene="mll1257"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48672.1"
/db_xref="GI:14022062"
/translation="MSSISARPIRSKSPVSATSRFMTRRTSSRAADRSGGPRDLLQKP
GFEPGFCLFKQVVFAYPRGMSALSRARHLLSGCAALALMLWLMPGSRADEPVDVELVL
AVDVSLSMSADELEIQRHGYAAALTHDNVLQAIADGAHGKIAVTYVEWAGTTWQRVIV
PWTVIANRADAERVVAQLSAQPPNSARRTSISGALEFGSDLFAESGYEGTKRVIDISG
DGPNNQGAPVNLTRDGVVRQGIVINGLPLMTRGGLSGAYDVNDLDRYYSDCVIGGPGA
FMIPVNDWTQFPEAIRRKLVLELAGPASPQWAAEEADHPPVVLTQDKPAADCQIGEKM
WRNRNWMLDSR"
gene complement(10906. .13467)
/gene="mll1258"
CDS complement(10906. .13467)

```
                 /gene="mll1258"
                 /codon_start=1

  Query Match             24.2%;  Score 247;  DB 1;  Length 329709;
  Best Local Similarity   56.3%;  Pred. No. 1.7e-25;
  Matches  504;  Conservative    0;  Mismatches  385;  Indels    6;  Gaps  2;

Qy      99 CGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGA 158
           || | ||   |  | |  | |     | | | ||| |||| |||||   |||| ||| |
Db 284688 CGGCGACCGCCACGTCGACCCGCATTCCTACCCCGACGGCATCGCCTTCCTCGACGGCCA 284747

Qy     159 ATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGA 218
           || ||||||      |||| ||| | ||  |  | |||   || |||   ||||||||
Db 284748 GTATCTGCCGATGTCGCAAGCCAAGGTGTCGGTGCTGGACTGGGGCTTCCTGCATTCCGA 284807

Qy     219 TCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCT 278
           ||||||  |   | |||||| ||| |||||  | |||||||| |||| |   || ||
Db 284808 CGCCACCTACGACACGGTGCATGTCTGGAACGGCCGCTTCTTCCGCCTCGACCTGCATCT 284867

Qy     279 GGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGA 338
           ||||| ||  ||| |||   |  |||||||| ||  |  |   | |  || ||| |
Db 284868 CGACCGCTTCTTCGGCGGACTGGAAAAGCTGCGCATGACCATCCCCTTCGACAGGGATGG 284927

Qy     339 ACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAA 398
           ||||||||||    | || || ||  |  | |||   |  ||||  | | ||  | |
Db 284928 CGTGGCCGAGATCCTGCACAATTGTGTCGCCCTTTCGGGCCATCGCGCCGCCTATGTCGA 284987

Qy     399 TCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCAC 458
           ||     ||| || ||   || |  ||  |     |       || | |  | ||   |
Db 284988 AATGCTGTGCACGCGCGG---CGCATCGCCGACCTTCAGCCGCGATCCGCGCCAGGCGAT 285044

Qy     459 CCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGAT 518
           | | |   |      || |||| | || | |   | ||   || |    |||||||| |
Db 285045 CAACCGCTTCATGGCCTTCGCCGTACCCTTCGGCTCGG---TCGCCAATGCCGAGCAGTT 285101

Qy     519 CTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGA 578
           ||  |       |||      || |       |||  |   |       | |||||
Db 285102 GCAGCGCGGCCTGCGCGTCGCCATCAGCGACAAGGTGCGCATCCCGCCGGCTTCGGTCGA 285161

Qy     579 CCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGA 638
           ||| | ||||||||||||||| |||   |||||       | |   | ||| ||| | ||
Db 285162 TCCGTCGATCAAGAACTACCATTGGCTCGATCTGGTGCGCGGCCTCTACGACGCCTATGA 285221

Qy     639 CCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGG 698
           | | ||||||   |||||| || | |||||||   ||  |   || |||| || ||||
Db 285222 CAGCGGTGCGGAGACCGCGCTCATTCTCGACTTCAACGGCAATGTCGCCGAGGGCCCGGG 285281

Qy     699 CTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCC 758
           |||||||| | |  |||||||||||||| ||| |  | ||| ||     || | ||||
Db 285282 CTTCAACGTCTTCTGCGTCAAGGACGGCAAACTGTCGACGCCGGCCATCGGCGTGCTGCC 285341

Qy     759 GGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCT 818
           |||||||| ||    || || ||||| ||   || | |    ||  | |  ||| ||
Db 285342 CGGCATCACGCGCCGCACAGTCTTCGATCTCTGCGCCGAAGAAGGTCTTGCCGCCGCCGC 285401

Qy     819 GCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGC 878
           |||  |||| |  |   |  ||| | |  |||||||||  |  |      | | |||||
Db 285402 CGCCGATGTCAGCGTCGCCGCGCTCAAGGCGGCCGACGAGGTCTTCATCACCTCGACCGC 285461

Qy     879 GGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGG 938
           |||||  |||  ||| | | |  | | || |||| |||  | | | ||||| ||  |||
Db 285462 CGGCGGCATCATGCCGGTGACCGAGATCGACGGCGCGGCGATCGCCGACGGCAAGGTCGG 285521

Qy     939 TCCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGC 993
           ||  | ||        |  ||  |  | ||||      |   ||||| |||| |
Db 285522 GCCGGTTACCAGCCGGCTAATGGCGCTCTACTGGCAAAAGCACGACGATCCGGCC 285576


RESULT 3
AP003015/c
LOCUS       AP003015               307551 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti plasmid pMLa DNA, complete genome, section 1/2.
ACCESSION   AP003015 BA000013
VERSION     AP003015.2  GI:14027841
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) plasmid:pMLa DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 307551)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994990.
FEATURES             Location/Qualifiers
     source          1. .307551
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
                     /plasmid="pMLa"
     gene            2138. .3256
                     /gene="mlr9002"
     CDS             2138. .3256
                     /gene="mlr9002"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB54434.1"
                     /db_xref="GI:14027842"
                     /translation="MEDYSEIFVGLDVSKDRHAVALAESGRSGEVRYVGEISSDSASV
                     RRLVRKLGRPGVRLRFCYEAGPTRYGLKREIEALGHECAVIAPSLIPRKPGDRVKTNR
                     RDAEKLARLFRAGELTEIWTPDAAHEAMRDLIRAREAAVKDRTRKRHEIRSFLLRHGK
                     IYPGLKAWGTKYLRWLHGLSFSYRSQEAVLQELILAESQCRERVVRLEKAIEDGLLDW
                     PLAPVVERLQALRGVQLICAATFMVEIGDVRRFANPRQLMAYLGLVPSERSTGDAVRR
                     GGITKTGNARVRRVLAESAWTYRYPAKLGKRKYCESRHMPEVVRDIAWNAQARLTKRY
                     RALIARGKRSTVAVTAIARELAAFMWAIARADEQQAGT"
     gene            complement(3881. .4138)
                     /gene="msl9003"
     CDS             complement(3881. .4138)
                     /gene="msl9003"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB54435.1"
                     /db_xref="GI:14027843"
                     /translation="MDEPLGVDPAQRMLANAELAGIIGYDDGIGEKPWWRIAPHIAPS
                     VAILTGFEDRALRRSAQQPARGTERATGLQANQSVVASRRW"
     gene            4469. .4591
                     /gene="msr9005"
     CDS             4469. .4591
                     /gene="msr9005"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB54436.1"
                     /db_xref="GI:14027844"
                     /translation="MKTCIKLDISFFRYLGGRLGIHAQPPIPPLPDLVRQAAQA"
     gene            complement(4625. .6148)
                     /gene="mll9006"
     CDS             complement(4625. .6148)
                     /gene="mll9006"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="probable oxidoreductase"
                     /protein_id="BAB54437.1"
                     /db_xref="GI:14027845"
                     /translation="MKGRFQPGNALQLEADAVIIGSGAGGASVADVLTAAGISVIMLE
                     EGKHIPSSSASPIASEAFLAAWRSGGLTAAIGRPPINYAEGRCVGGGTEINSAITQRA
                     DSNLIDQWRELYKIENFTPDELAQYYERAETAVNASLTPGPLGRATDILRKGGEALGW
                     KVSELKRGQRGCKGANRCSYVCPNGAKQSMAITLLPKAVDRGMRLLTQTRVDNIRIEK
                     GRATEVLARSRDPGGHAMAVRVKAGMVFVCAGAIHTPALLRRSGPRKCIGNTLRVHPT
                     IRATALFDEPVDAHQSRLPLTAITEFMPEQRIGGSVFTPAIFGLSLAEDWTNRGDLMR
                     NWRFCGSYYGMIRPRGVGSVRPLPGINEPLVGFKLAPDDWIALCQVLTRLGQVMFEAG
                     ARKVFPGISGHEGWTSPNQVHEYWNKPLPKQATNLMSIHLFSSCPPGEHRDACAVDSY
                     GRVRGVENLFVADGSLIPEAPGVNPQMTIMALAFRAAEAALCHSAQERARSAARDRN"
gene                 complement(6534. .6887)
                     /gene="mll9408"
CDS                  complement(6534. .6887)
                     /gene="mll9408"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB54438.1"
                     /db_xref="GI:14027846"
                     /translation="MNTPPVIMVLAAGLNSCIGNMLLKWSRISLPTDAGPTDKFLSIS
                     FIGGLLFYGINVVFFAKALDSMEVSVAYPILAGSAFAMLAVASYYIFGEPFHLSKRIG
                     LSLVLAGIILLARGD"
gene                 complement(6884. .7933)
                     /gene="mll9007"
CDS                  complement(6884. .7933)
                     /gene="mll9007"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB54439.1"
                     /db_xref="GI:14027847"
                     /translation="MVNTALVTGGSGYFGELLTKQLLEQGTHVRVFDLNFPGFSHPNL
                     EFFKGTILDRNAVKQAVSGVKKVFHNVAQVPLAKDVDLFWSINKDGTQILADESVAAG
                     VVKLIYTSSSAVFGVPKANPVTENTVTNPAEDYGRAKLAGEMICKEAMQRHGLDVAIV
                     RPRTILGHGRLGIVQILFDWIERGLDIPVFAGGNNRYQFVHSDDLAWACIAASDMRGF
                     AEYNIGAAEFGTMRELLQTVIDHTGTESRIKSIPMRPTALAANLASTLGLSPLGPYHS
                     LMYGRSMYFDISKAQRELDFMPKYSNSQMIIETYDWYQANKASLSKGRASRHKSPVKQ
                     QILALVPYVLRMAPG"
gene                 complement(9609. .10418)
                     /gene="mll9009"
CDS                  complement(9609. .10418)
                     /gene="mll9009"
                     /codon_start=1
                     /transl_table=11
                     /product="transcription regulator"
                     /protein_id="BAB54440.1"
                     /db_xref="GI:14027848"
                     /translation="MHVKTNEKKEKATGTQLLDRSIAILNHLGELGRGGARVSEIAEA
                     FGLNMSTAHRIISGLERHRYIERDDTTKRYRLGLALFALGAKAADTTGMRRVCHPALL
                     RIAAETGDTVFLMARNGFNTVCVDRQDGGYRIDSLTGQVGGQIPLGVGPASQAILAML
                     PPNQAEVILETNAPQFERYHGLTADEIRHALPEIRGRGYAVDDGHLVEGISAIAVPII
                     LESGNVAGALAINMTSARLRRCSVGDFVQILRREIDEIERTMNPHDVSIAS"
gene                 complement(10505. .11878)
                     /gene="mll9010"
CDS                  complement(10505. .11878)
                     /gene="mll9010"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB54441.1"
                     /db_xref="GI:14027849"
                     /translation="MTIARHFADFLSEVREVNDLAVEQVKAAVLDLATAALAGVNEDA
                     TVAARNAALMIWQHGNAPIWFTKHRTSAAGAAFVNSTSASSLDLDDGHRAAAGHPGAS
                     IIPAVMATAAEFKCPTNRILPAIALGYEIAVRVAASRDFGTLTTLVSGPWVGYGAAAA
                     AAWLRGLEPRAIEQAIAIAGSTAPNLSAIAYSEVMGNHVKEGIPWATAAALAAVELAA
                     QGFTGPDDLLDNERVFDRGRLLDGLGTEWAISGIYFKPYGCCRWAHSAIDAVLALKAE
                     HGLHATEIQAIEIETFSWALRLNNEVAPRTTESAQYSVPFCVAMALTKGAEALLQITN
                     RDLADSDTIALARRIQLRVDRGFDATFPQSVPARVIMRTKNERLSKEVLAPLGEPTNP
                     MGWATLENKFRNVARPRLDEQHLEMFLVAFRDLRDGELGSLMETLQRPIPRFVSIDEP
                     ALEANCQ"
gene                 12307. .13443
                     /gene="mlr9012"
CDS                  12307. .13443
                     /gene="mlr9012"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical hydrolase/peptidase"
                     /protein_id="BAB54442.1"
                     /db_xref="GI:14027850"
                     /translation="MTANRQAFTTEEFKTRVARVQSEMSVRGIDILLLHAPENIYYLT
                     GHQTSGYFAYQTVVLAQAGDPQLLLRFLEKGNVDEYSWLPDASTWKEGDDVVEKTLAL
                     VRGFGAAGKRIGLEKRSWFLTSAIADALTNGLSDATIVDASLLVDRVRVIKSPAEVGY
                     IRRAAQIAEIEQQVAFATICDGATEAQVAAAVFAAGVNAGCEYTGLPHHIMSGYRYNV
                     CHANWSPKIIRRGELVLFEFYGCLERYHATQMRTVAIGHASDEVTRAADIVVAAQDAG
                     IAAMKPGASAREVDALVRQPIRKIRPEYYNRTGYSTGIGFPPKTAEWEALDFNEQEDW
                     ELKEGMAFHMLALACGFGISETVVVTSNGVERLTGSNGRALLIK"
gene                 13484. .14587
                     /gene="mlr9013"
CDS                  13484. .14587
                     /gene="mlr9013"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB54443.1"
                     /db_xref="GI:14027851"
                     /translation="MTDKIFNIRVDRRSFLAGSAAIAASTLPSPLIRSASAQEFAGKQ
                     IRVLTWSDPTGQAAVKNIMQTFEAKTGAKVIADLTGSTSDMIAKIKASASKPQYDVVV
                     LSGFGADSLAEAGLLEKPDTSAIANLSRVDPKYQTGANGHGIGYFLWSDGLIYNTNRY
                     TTPPKSYDVIWNAANEGKIVLPTPDNLGALELIIVAAKAAGGDLKNPDPGFELLKKMK
                     GKALAVTTNSNQIADLFRSNAVNAAGVMSPLTFSTFIPNPEFNIGGTYDLAEGFFVDL
                     QYMVIPKGHPGDAATVNAFLNHALDPVVQGKMAEEVWYGPINQDAILSDKAKQSPFIP
                     SPAVIKERANRLDSGFLASVRDDWIQRYTDAVM"
gene                 14669. .15748
                     /gene="mlr9014"
CDS                  14669. .15748
                     /gene="mlr9014"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, ATP-binding protein"
                     /protein_id="BAB54444.1"
                     /db_xref="GI:14027852"
                     /translation="MEQNPVLEIRNLSRTFAHGAVRAVDHIDLDVLPGEIMSILGPSG
                     CGKTTIMRIIAGLDSPTSGEVRILGKDVAGKPPHKRNVGLVFQSLAIFPHMSVKQNVA
                     FGLRMKGLASNLIETKVAKALELVQLSPERFAGRRPSELSGGQLQRVALARTLVTEPA
                     LVLFDEPMAALDRRLRDYMAVELRAIQKQLGIAAVYVTHDQETASMMSDRVIIMSAGK
                     ILQAGTPEEIYEAPNSRFVSDFLGDTNTLRIARVLHQADGTSLVQTDAGLAIRVANCS
                     DIAEGAFVIFRPEQTIVHSADPGGAFEGRVVSSQFRSGLHRWQIALPGGNTLIAQSTE
                     NRLASSAGETAWVTIDSGKARIVSQ"
gene                 15745. .16638
                     /gene="mlr9015"
```

Query Match 18.8%; Score 191.6; DB 1; Length 307551;
Best Local Similarity 52.9%; Pred. No. 6.7e-18;
Matches 460; Conservative 0; Mismatches 404; Indels 6; Gaps 2;

```
Qy    122 GCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGA 181
          | | ||   |   |||||| || |  || || ||  | |||||||||     |||||
Db 184870 GACAGTATCCTCACGGCGTGGCGTTCATGGATGGTCAGTACCTGCCGATGTCCGAAGCCA 184811

Qy    182 AGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATG 241
          || | ||  |  | |||   || ||    |||||||||   ||||||     ||  ||||
Db 184810 AGGTGTCGGTTCTGGACTGGGGCTTTCTGCATTCCGATGCCACCTACGATACCGTTCATG 184751

Qy    242 TATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGT 301
          | ||| |||| |   ||||  ||||       ||| |||||||||  ||  ||||  |
Db 184750 TGTGGGACGGGAGGTTCTTTAGGCTGAATCTGCACGTGGACCGGTTCTTCCGCGGGATGG 184691

Qy    302 CCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGT 361
            ||||| ||| || |       |||| |       ||  | |   |||||     ||  |
Db 184690 AAAAGCTCCGCATGAAGCTTCCCTACAACCGCAGTGAGATCGAAAAGATCCTGTCGACCT 184631

Qy    362 GCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACG 421
```

```
Db 184630 GTGTAGCGCTCTCCGGACACAAGTCGGCCTATGTCGAAATGATCTGTACGCGAGG---TG 184574

Qy    422 GAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCA 481
Db 184573 GCTCGCCCACCTTTAGCCGGGACCCGCGCCAGTCGGAGAACCGGTTCATCGCATTCGCGG 184514

Qy    482 TCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGC 541
Db 184513 TGCCTTTCGGGTCGG---TTGCCAACAAGGAGCAGCTCGAGCGCGGCCTGCATGTCGCCA 184457

Qy    542 CGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGT 601
Db 184456 TCAGCAACACTGTCCGGATCCCGCCCAAGTCGATCGATCCGACCATCAAGAACTATCACT 184397

Qy    602 GGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCC 661
Db 184396 GGCTCGATCTGGTGAAGGGCCTGTTCGATGCGTACGACTATGGTGCCGAAACTGCATTGA 184337

Qy    662 TGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGG 721
Db 184336 TCGTCGACATCAACGACAACATTGCCGAAGGGCCTGGCTTTAACGTTTTCACCGTCAAAG 184277

Qy    722 ACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGT 781
Db 184276 ACGGCCGCTTGAAGACGCCGGCTTATGGTGTCCTTGCCGGCATTACCAGGCAGACCGTAT 184217

Qy    782 TCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAAC 841
Db 184216 TTGATCTGTGCGACGAACTGGGCCTATCCGTCAGCGCCGGCGACATTGATCGCAACGAAC 184157

Qy    842 TCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACT 901
Db 184156 TGAAGGGCGCCGATGAGGTTTTCATCACGTCTACAGCAGGCGGGATCATGCCGGTGTCCA 184097

Qy    902 CGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGG 961
Db 184096 AGATTGACGAAACCGTTGTAGGCGACGGCAAGGTTGGCGCTCTAACACGCCAGCTGGCAG 184037

Qy    962 ACCGGTTCTGGGCGCTGATGGACGAGCCGG 991
Db 184036 ACCTGTATTGGGAGAAGCATGCTGATCCCG 184007
```

RESULT 4
MLO672114/c
LOCUS MLO672114 149050 bp DNA linear BCT 14-MAY-2002
DEFINITION Mesorhizobium loti R7A symbiosis island; segment 3/4.
ACCESSION AL672114 AL672111
VERSION AL672114.1 GI:20804067
KEYWORDS .
SOURCE Mesorhizobium loti.
ORGANISM Mesorhizobium loti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE 1 (bases 1 to 149050)
AUTHORS Sullivan,J.T., Trzebiatowski,J.R., Cruickshank,R.W., Gouzy,J.,
Brown,S.D., Elliot,R.M., Fleetwood,D.J., McCallum,N.G.,
Rossbach,U., Stuart,G.S., Weaver,J.E., Webby,R.J., de Bruijn,F.J.
and Ronson,C.W.
TITLE Comparative sequence analysis of the symbiosis island of
Mesorhizobium loti strain R7A
JOURNAL J. Bacteriol. 184 (11), 3086-3095 (2002)
MEDLINE 21999272
PUBMED 12003951
REFERENCE 2 (bases 1 to 149050)
AUTHORS Ronson,C.W.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2002)
COMMENT Department of Microbiology, University of Otago, P.O. Box 56,
Dunedin, New Zealand; MSU-DOE Plant Research Laboratory, Michigan
State University, East Lansing, Michigan 48824 USA; Laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes, Unite
Mixte de Recherche (UMR) 215 Centre National de la Recherche
Scientifique (CNRS), Institut National de la Recherche Agronomique,
Chemin de Borde Rouge, BP27, F-31326 Castanet Tolosan Cedex,
France. clive.ronson@stonebow.otago.ac.nz
http://sequence.toulouse.inra.fr/msi.

FEATURES Location/Qualifiers
source 1. .149050
/organism="Mesorhizobium loti"
/strain="R7A"
/db_xref="taxon:381"
repeat_region 454. .1516
/note="predicted by Homology"
/insertion_seq="msi437"
gene 520. .786
/gene="msi239"
CDS 520. .786
/gene="msi239"
/function="elements of external origin; transposon-related functions"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSPOSASE FOR INSERTION ELEMENT PROTEIN"
/protein_id="CAD31271.1"
/db_xref="GI:20804068"
/translation="MKRKRFTEEQIIAVLREHEAGAKAGDLARKHAISEATLYNWKAK
YGGMDVSDAKRLKALEDENAKLKKLLADQMLEAAALKELLSKKW"
gene 780. .1709
/gene="msi240"
CDS 780. .1709
/gene="msi240"
/function="elements of external origin; transposon-related functions"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE"
/protein_id="CAD31272.1"
/db_xref="GI:20804069"
/translation="MVGPAARREAVAHLQAVMGLSERRACSIVGADRKMVRYRSCRPP
ETELRGRLRELANERRRLGYRRLFILLRREGEASGVNRIYRLYREEGLTVRKRRARRR
AVGTRAPILVEARPNARWSLDFVHDQFACGRRFRVMNIVDDVTRECLAAIPDTSISGR
RVARELTDLISLRGKPDMIVSDHGTEFTSNAILAWSKDHRVEWHYIAPGKPMQNGYIE
SFNGRMRDELLNESLFFGLDHARCAITEWVEDYNTARPHSSLGYQTPAAFAEVLTATG
SNAALDVGFASPPVAQPAPYGVTETVEALIAAG"
gene complement(1825. .2691)
/gene="msi241"
CDS complement(1825. .2691)
/gene="msi241"
/function="cell processes; transport of small molecules"
/note="Product confidence : putative
Gene name confidence : putative"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE AMINO-ACID ABC TRANSPORTER ATP-BINDING PROTEIN"
/protein_id="CAD31273.1"
/db_xref="GI:20804070"
/translation="MTATDIDHLIGLNGGDSSMATGNVRRIPVSSQALSGQKQVPMSA
DKIIEATHVSKWYGSFRALTDVNLTVRKGERIVICGPSGSGKSTLIRCFNRLEANQEG
EIRVNGMLLHNKMRNVTEVRKNVGMVFQHFNLFPHMTVLMNCMAGPMWIKGVPEAQAK
QTALKYLERVRIPEQANKYPVQLSGGQQQRVAIARSLCMEPAVMLFDEPTSALDPEMV
SEVLETMTSLARDGMTMVCVTHEMGFARAVADRIIFMDKGRIVEEANPNDFFTNPHHE
RTRLFLSQILSH"
gene complement(2688. .3569)
/gene="msi242"
CDS complement(2688. .3569)

```
                /gene="msi242"
                /function="cell processes; transport of small molecules"
                /note="Product confidence : putative
                Gene name confidence : hypothetical"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PUTATIVE AMINO-ACID ABC TRANSPORTER PERMEASE
                PROTEIN"
                /protein_id="CAD31274.1"
                /db_xref="GI:20804071"
                /translation="MFNLSGTAFGDLMRPVIGDPLQSGIYGRVAIAFVIAVMFCVNIV
                LIGFAPLRVQIGIVWAELLLLFLGFFNAFNLNMPFIWEHLPYLITQGVVTTIYVSAVS
                LIFASIIAMIAAIAKLSSNGFAYALASFYTSFFRGLPLLIQIYLIYLGLPQLGIVINA
                VPSGILALSLCVGAYMTEIFRAGIQSIDRGQWEASRSIGFGFGLTIRKIILPQALPVI
                IPPMGNTFIAMLKDSSLVSVLGVWELTFLARTIGQPTFHHMEMLITAAMIYWILSACL
                EVLQSRLEHHFARSKVR"
gene            complement(3724. .4557)
                /gene="msi243"
CDS             complement(3724. .4557)
                /gene="msi243"
                /function="cell processes; transport of small molecules"
                /note="Product confidence : putative
                Gene name confidence : hypothetical"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PUTATIVE AMINO ACID ABC TRANSPORTER PERIPLASMIC
                BINDING PROTEIN"
                /protein_id="CAD31275.1"
                /db_xref="GI:20804072"
                /translation="MKKIKMAVTAVLIALTSVAFGPSAHAGEVLDRILTSKTLRVAAG
                ADWGLIAHLDEKNELVGYDIDVAKGIAKYLGVNAEFVTPGWDLIAAGNWAGRWDIGTG
                QMAPTKARADKFDFSVPYFYDRTVIVVHKDSKATKGSDLEGKKIGVNGGTVAEAYANQ
                TYTPDDPNAAPVPYDFKAGEVKTYGNSTIAFDDLRLGDGVRLDAVITDETFTEGAIKS
                GYPIKIIDRLFALPAAIPMMKGDKELVEKVSAAIKSMRDDGTLSKLSVKWYGSDHSVA
                K"
gene            complement(4612. .4797)
                /gene="msi244"
CDS             complement(4612. .4797)
                /gene="msi244"
                /function="miscellaneous; unknown"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="HYPOTHETICAL PROTEIN"
                /protein_id="CAD31276.1"
                /db_xref="GI:20804073"
                /translation="MTTANESLLGLETKTSGAMCERLDTVRSHLAWDLVSIGSHWLSV
                YPHDSETSYHEEAGSEI"
repeat_region   5015. .6684
                /note="predicted by Homology"
                /insertion_seq="msi442"
gene            5155. .6489
                /gene="msi245"
CDS             5155. .6489
                /gene="msi245"
                /function="elements of external origin; transposon-related
                functions"
                /note="Product confidence : putative
                Gene name confidence : hypothetical"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PUTATIVE TRANSPOSASE FOR INSERTION ELEMENT
                PROTEIN"
                /protein_id="CAD31277.1"
                /db_xref="GI:20804074"
                /translation="MRQERTVQSSVFDLFAGHEIGCELKAMSQWLDGHRELLSLVTRD
                LRRHDLKETGREGLPAEAVLRCALLKQHRQLSYQELAFHLEDSASFRAFARLPWGWSP
                KKSVLHKTISAIQAPTWEEINKVLLASAGREKLEGGKVVRVDSTVTAALVHEPSDSSL
                LWDAVRVMVRLLQHADALGSAIPWHDHCRAAKKRSRSIEYTRGRPKRVKHYRELLRIT
                RTTLGYLQQAAEQLTLAACPAVELWQAQVRHYRPLIERIIAQTERRVLAGEPVPAGEK
                LVSLFEPHADIIVKGSRDVDYGHKINLTTGRSGLILDLVIEAGNPADSERLLLMLERH
                IAFYGEPPRQAAADGGYASRPNLRQAKAWGVRDMAFHKKSALRIEDMVSSRWVYRKLR
                NFRAGIEASISCLKRAYGLARCTWRGLNHFKAYVWSSVVAYNLALFTRLRSN"
misc_feature    6804. .6995
                /note="REGION CONTAINS INSERTION SEQUENCE FRAGMENT
                predicted by Homology"
                /evidence=not_experimental
gene            complement(7356. .8585)
                /gene="kbl"
CDS             complement(7356. .8585)
                /gene="kbl"
                /EC_number="2.3.1.29"
                /function="small molecule metabolism"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="HYPOTHETICAL 2-AMINO-3-KETOBUTYRATE COENZYME A
                LIGASE 2.3.1.29 PROTEIN"
                /protein_id="CAD31278.1"
                /db_xref="GI:20804075"
                /translation="MTQTFLAHLRRELADLKSGGLYKSERVISSRQSREIRVANGNSV
                LNFCANNYLGLADNEEIADAAKKALDRYGFGMASVRFICGTQEEHKQLEARISGFLGM
                EDTILYSSCFDANGGLFETLLGEEDAIISDALNHASIIDGARLSKARRFRYANNDMKA
```

Query Match 15.0%; Score 153.2; DB 1; Length 149050;
Best Local Similarity 50.4%; Pred. No. 1.5e-12;
Matches 430; Conservative 0; Mismatches 418; Indels 6; Gaps 2;

```
Qy    119 CCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAG 178
Db   9550 CCCGGCGTTATCCTCACGGTGTCGCCTTCATGGACGGACAGTACCTCCCAATGTCTGAAG 9491

Qy    179 CGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGC 238
Db   9490 CGAAGATCTCTGTCCTTGACTGGGGTTTCCTGCACTCTGACGCTACTTACGATACCGTGC 9431

Qy    239 ATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGG 298
Db   9430 ACGTCTGGGAAGGGCGATTCTTCAGACTGGACCTTCATCTTGATCGCTTCTTCAGGGGGA 9371

Qy    299 CGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGA 358
Db   9370 TGGATCGTCTCCGAATGAAGCTGCCCTATCATCGCCGAGAGGTCGAACGGGTTCTTTCAA 9311

Qy    359 AGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGAT 418
Db   9310 ACTGCGTCGCTCTGTCTGGACATAAATCAGCCTATGTTGAGATGATCTGCACGCGGGGA- 9252

Qy    419 ACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACG 478
Db   9251 --GGTTCGCCGACATTTAGCCGCGATCCGCGCGAGGCCGAGAATCGGTTCATTGCCTTTG 9194

Qy    479 CCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCG 538
Db   9193 CTGTGCCGTTTGGTTCGGTTGCCAACAAAGAACAGTTGGAGCGTGGTCTCCACGTTGGCG 9134

Qy    539 TGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACC 598
Db   9133 TAAGCGAAACTGTCCGAATTCCG---CCGAAGTCAGTGGATCCGACGATCAAGAATTACC 9077

Qy    599 AGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGA 658
Db   9076 ACTGGCTTGACTTGGTCAGAGGTCTTTACGATGCTTACGATGTGGGGGCGGAAACGGCGC 9017

Qy    659 TCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCA 718
Db   9016 TCATAATGGACACCAACGGCAATATTGCGGAAGGTCCCGGCTTTAATGTCTTCACCGTTA 8957

Qy    719 AGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGG 778
```

```
          || || |  |||||    | ||  ||      | | | || || || || ||  || | |
Db    8956 AGAACCGTCAGCTGAAGACACCAGCCTTTGGGGTGCTTCCAGGAATAACGCGGCAGTCCG 8897

Qy     779 TGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTG 838
          | ||||| |||   |   |  ||||  | |  |  ||       ||| |  |  | | ||
Db    8896 TTTTCGACCTGTGTGGGGAAGTGGGACTTGCGGTGACGGCTGCAGACCTACCTCGTCTTG 8837

Qy     839 AACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCA 898
          ||||   ||| ||||| ||| || |      | |||||||| || ||| |||  ||  | |
Db    8836 AACTGGGCGAAGCCGATGAGGTGTTCATCACCTCCACCGCAGGGGGGATCATGCCCGTGA 8777

Qy     899 ACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCC 958
           | |||||     || | |||| |||  |    ||     ||||   |    |
Db    8776 CGCGCGTTGATGGTTCATCCATCGGCAGCGGTAAAGTTGGCGTCGTGACCCGGCAGCTGA 8717

Qy     959 GGGACCGGTTCTGG 972
           |||||  | ||||
Db    8716 TGGACCTCTACTGG 8703


RESULT 5
AE005957/c
LOCUS       AE005957               10364 bp    DNA     linear   BCT 12-JUN-2002
DEFINITION  Caulobacter crescentus CB15 section 283 of 359 of the complete
            genome.
ACCESSION   AE005957 AE005673
VERSION     AE005957.1  GI:13424548
KEYWORDS    .
SOURCE      Caulobacter crescentus CB15.
  ORGANISM  Caulobacter crescentus CB15
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 10364)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
            Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
            Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
            Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
            Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
            Fraser,C.M.
  TITLE     Complete genome sequence of Caulobacter crescentus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
  MEDLINE   21173698
   PUBMED   11259647
REFERENCE   2  (bases 1 to 10364)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
            Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
            Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
            Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
            Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10364
                     /organism="Caulobacter crescentus CB15"
                     /strain="CB15"
                     /db_xref="taxon:190650"
     gene            69. .1196
                     /gene="CC2925"
     CDS             69. .1196
                     /gene="CC2925"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK24887.1"
                     /db_xref="GI:13424549"
                     /translation="MMLTLLAALAVAAAPCPASDESPACVRQRIDALPMTDLLTVGTH
                     NSYKLAIPPEEMAAMVAARGPVALGIDYSHRPLTEQLDAGARQLEIDVVADPQGGLFA
                     KPLTALGKGVTLTPDVAAALAKPGFKTLHIPDVDFRSSCVTFVACLKEVRAWSDTHRD
                     HAPILIMLNAKEGAASLPGGVTPLPFTEALFDALDAEIRAVFGEDRLITPDQVQGKAK
                     TLRDGVLAGGWPKLGAARGKVFFALDESPAKVALYRGQRASLEGRAMFVNTDEASPAA
                     AYLTLNDPVAQKDRIAAAVKAGFIVRTRADAETLQARRNDPTQRDAALASGAQYVSTD
                     YIWADPRFAGDYTVRLPGNKVAVCNPVRQPTGCGGRDLEGR"
     gene            complement(1206. .1418)
                     /gene="CC2926"
     CDS             complement(1206. .1418)
                     /gene="CC2926"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK24888.1"
                     /db_xref="GI:13424550"
                     /translation="MAGQAEGRAMKTGHHLTLWQVFGAPALIAAFSLFGLIAALLGDG
                     LWDLIGVLTLGVSVIATVWALAARRR"
     gene            complement(1388. .2572)
                     /gene="CC2927"
     CDS             complement(1388. .2572)
                     /gene="CC2927"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK24889.1"
                     /db_xref="GI:13424551"
                     /translation="MGGGRRLSGRQLPDAGRAAHPAPVDQHGLLAPMKARTLRAWSWV
                     HKWSSLVSTAFLLMLCITGLPLVFSHEIEHVLMEEAWTPARPDGPKLNLDQVLDTALA
                     RKPGEVPAYMSFDEDRPVVNVTSVQPGGKAGVYSFQPIDQTSGDPAPPVAGHPVMEFL
                     LQLHTDMFLGLAGMLFLGAMGLLLIVALVSGVVLYAPFMRRLPFGTVRASKRARTRWL
                     DYHNLLGAVTAAWVLVVGATGVVNALATPIVGYWKNTALKELTAAYDSPVAPGQRSSL
                     QAAVDKAKQALPDKDLQFVAFPGSDYSTDHHYAVFFHGKTPLTTHLTTPALIDARTGE
                     LAAIAPTPWYVKTLTLSQPLHFGDYGGMGLKILWALLDIATIVILASGLYLWLAKRKV
                     AR"
     gene            complement(2482. .4647)
                     /gene="CC2928"
     CDS             complement(2482. .4647)
                     /gene="CC2928"
                     /note="identified by match to  protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="TonB-dependent receptor"
                     /protein_id="AAK24890.1"
                     /db_xref="GI:13424552"
                     /translation="MSSRLALLLSSALISGALASPALARQADTTRVDEVVITGSQVRL
                     PPAYAGDQVAKGARVGLLGALDTMDTPFAVTSYTEALIRNQQARSVADVLQNDPTVRV
                     TKGFGNFQELYVVRGFPVYSDDMTYNGLYGVLPRQFVAAELIERVDVFRGANAFLNGA
                     APGGSGVGGAFNLMPKRAPTAPLTRFTAGWDGGEQVYGSADIARRFGQNDAYGARVNL
                     GLRGGETAVDGEKRDLTVLGLGLDRRGDRARFSADLGWQDHRIEGARPTVTPLGAIPK
                     APSADKNFAQPWTHTDERQLFGAARGEFDLTDTVSAWAALGGRQGKEDNVLANPTALA
                     NGTLSAYRFDNVRKDSILSTDAGLRAKFTTGAVGHALVASVAQVNLKSKNAYAFSNFA
                     GFASNLYNPVAVAAPNPNFFVGGSKSDPNVTERVKNRSVAVADTLSFLNERLLVTVGV
                     RYQDIQTRSYDYNTGARNGAYDGDATTPALAAVFKPSDKISLYANYAEALVPGKIAPA
                     QVNGVAVANVGEILSPFRGEQTEIGAKYDAGTFGGSISLFRTTQPAEYFEASSRRYTA
                     GGEQENQGVELTVFGQPAPGLRLLGGATWLDAQINRALSATLQGKTPIGVPEFQANLN
                     LEWDVATLEGLTLEGRAVHTGSQQANTANTVSLDAWTRFDLGARYAFEAGGKAVTLRA
                     RVENLTDRNQWVAVGGYPGANYLTLGAPRTLRLSISTDF"
     gene            complement(4868. .5758)
                     /gene="CC2929"
     CDS             complement(4868. .5758)
                     /gene="CC2929"
                     /note="identified by match to PFAM  protein family HMM
                     PF01063"
                     /codon_start=1
                     /transl_table=11
                     /product="branched-chain amino acid aminotransferase"
                     /protein_id="AAK24891.1"
                     /db_xref="GI:13424553"
                     /translation="MSLVPFDDRDGWIWLDGQFVPWREAKVHVLTHGLHYASSVFEGE
                     RMYGGEIFKLTEHTERLFKSAEILDFEIPYTVAEIDEACKATAAKNGLKDCYVRPIAW
                     RGSEMIGVSAQQSKIHVAIAVWEWPSYFDPATKAKGIRLTWAKYQRPDPKTAPIAAKA
```

```
                AGLYMICTISKHAAEKDGYADAMMLDYRGYVAEATGANVFFVKDGVLHTPKPDCFLDG
                ITRRTVIDLAKAKGIEVVERHIQKEELATFTECFIVGTAAEVTPVSEVGEFKFTPAKL
                SLDLMDTYAALVRGEAMATA"
gene            5839. .6345
                /gene="CC2930"
CDS             5839. .6345
                /gene="CC2930"
                /note="identified by match to  protein family HMM"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator, MarR family"
                /protein_id="AAK24892.1"
                /db_xref="GI:13424554"
                /translation="MIAPLQPGSDDPRLILREEELDGGLELILLAEASLWAAVDAVLE
                TEALGLGRSHWRAAFLLRRRPGIGVQDLSKLTSLSKQAASRTLSDLEKAGLVERVSGD
                LDGRRRPATLTAEGVAFEQRTAERLRALLARAYRTGGLDGVAGTRRILAALAGSRQGV
                GPGRRMPT"
gene            6342. .7043
                /gene="CC2931"
CDS             6342. .7043
                /gene="CC2931"
                /note="identified by match to  protein family HMM"
                /codon_start=1
                /transl_table=11
                /product="petR protein, putative"
                /protein_id="AAK24893.1"
                /db_xref="GI:13424555"
                /translation="MTPDERRERHLLVVDDDDRLRKLIKEFLSRAGFRVTAASSAAAA
                DKLFDALDFDLMVLDVMMPGEDGMAFTKRLRAKGGEAGRTPILMLTARDQTADRIEGL
                SSGVDDYLGKPFEPQELLLRIEAILRRSAARPVGPKALSLGRCTFDADRGELTCDGEA
                VRITEAEVTLLRRLARSLHEPVDRLELARDTADATGRAVDVQVTRLRRKIEPDPKNPR
                YLQTVRGVGYRLAPD"
gene            7206. .8543
                /gene="CC2932"
CDS             7206. .8543
                /gene="CC2932"
                /note="identified by match to  protein family HMM"
                /codon_start=1
                /transl_table=11
                /product="sensor histidine kinase"
                /protein_id="AAK24894.1"
                /db_xref="GI:13424556"
                /translation="MPLTRLDIPPALKRLLPTTLFGRSLLIIILPVAIMQIAVTWAFF
                DAHWQSVTSKLSEGLAGDIAWAVQSYEDDPSPAAVEKLAERAEGALSLSIAFQKGRKL
                PTSHRPSLFAALDRSLDKALEDRLDNPFWFDTTRYRAYIDIRVQVDGGVLQIYALRDR
                AYATQGHIFILWMVVATLLLTAVAILFIRNQVRAIERLAEAADAFGRGEDPEFKPHGA
                REVRQAALAFIAMKLRIQRHIEQRTALLASVSHDLRTPLTRLKLEMAMAEPCERMEAM
                KGDLAEMEHMIDEYLAFARGEGGEAIQVVDLSDLVESVVADAERGGAAIETEITQGLE
                TRLRPLTFRRALANLIDNGVAHADRVRVTVSPRQTGGVDVAVDDDGPGIPEDKYEEAF
                KPFSRLDESRNQNEKGVGLGLAIARDMARGLGGDLVLSRSALGGLRALIRLPG"
gene            complement(8671. .9522)
                /gene="CC2933"
CDS             complement(8671. .9522)
                /gene="CC2933"
                /note="identified by match to PFAM  protein family HMM
                PF00800"
                /codon_start=1
                /transl_table=11
                /product="prephenate dehydratase"
                /protein_id="AAK24895.1"
                /db_xref="GI:13424557"
                /translation="MSLLKKIAFQGEPGANSHEACRTYFPDYEAYPCKTFEEAFEAIK
                SGVAQLGMIPIENSIAGRVADVHHLLPASGLKIIGERFKPIRFQLMANKGVKLEDIKV
                VSSMPIALSQCRNSLKRLGVETEAAGDTAGAAKALALKPNPTHAAVAPALAAEIYGLD
                ILARDIEDERHNTTRFLVMTADKAPAAPDFTHRCVTSFVFRVRNLPAALYKALGGFAT
                NGVNMTKLESYMEGGNFTATFFYAEVDGRPEDRNLALALDELKFFSERFEILGVYPAD
                PFRDRGA"
gene            complement(9519. .10256)
                /gene="CC2934"
CDS             complement(9519. .10256)
                /gene="CC2934"
                /note="identified by match to TIGR  protein family HMM
                TIGR00466"
                /codon_start=1
                /transl_table=11
                /product="3-deoxy-manno-octulosonate cytidylyltransferase"

 Query Match           12.7%;  Score 129.4;  DB 1;  Length 10364;
 Best Local Similarity  49.9%;  Pred. No. 4.9e-09;
 Matches  384;  Conservative    0;  Mismatches  376;  Indels    9;  Gaps    2;

Qy    132 CGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCAT 191
          || |||       |||| | || ||  |||   ||||        |||||||||  |     |
Db   5732 CGACGGTTGGATCTGGCTTGACGGGCAATTTGTGCCCTGGCGCGAAGCGAAGGTGCATGT 5673

Qy    192 CTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGG 251
          ||   |  ||  |      || |     |   | |   | || ||   | |    |||
Db   5672 TTTGACCCATGGCCTGCACTATGCGTCGTCGGTGTTCGAAGGCGAGCGCATGTACGGCGG 5613

Qy    252 CAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCG 311
          | | ||||||   ||   ||| ||    || ||  || || |   |||    |  ||
Db   5612 CGAAATCTTCAAACTGACCGAGCATACCGAGCGCCTGTTCAAGTCGGCCGAGATCCTCGA 5553

Qy    312 CCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGAT 371
          | ||   |  ||||    || |||  | | ||||  |  ||||  |   | | | |
Db   5552 TTTTGAGATCCCCTACACGGTGGCCGAGATCGACGAGGCCTGCAAGGCGACGGCGGCCAA 5493

Qy    372 GTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAA 431
          |     |||   || |  |  |||  || || || |   || ||   || || |  |
Db   5492 GAACGGTCTGAAGGACTGCTATGTGCGTCCGATCGCCTGGCGCGGCAGCGAAATGATCGG 5433

Qy    432 GGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCT 491
          |    |  | | | ||||||| ||      || |||  ||||  || |     ||  |
Db   5432 TGTTTCGGCGCAGCAGTCCAAGATC------CACGTGGCCATCGCCGTCTGGGAGTGGCC 5379

Qy    492 GTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGT 551
          |   ||||   || |||    ||  |   ||||  |  |  || |  | |     |
Db   5378 CAGCTACTTCGATCCGGCCACCAAGGCCAAGGGCATCCGCCTGACCTGGGCCAAGTACCA 5319

Qy    552 CCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCT 611
          || | ||  |  |   | | ||     | || |||       |  ||  |
Db   5318 GCGTCCCGACCCCAAGACCGCCCCGATCGCGGCCAAGGCCGCCGGCCTCTACATGATCTG 5259

Qy    612 CACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTC 671
          ||||   |||    ||  ||| || |    ||      |  ||||||  |||| ||||
Db   5258 CACCATCTCCAAGCACGCCGCCGAGAAGGACGGCTATGCCGACGCGATGATGCTGGACTA 5199

Qy    672 GGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCT 731
          |  ||  || ||||| |   | |||  |||||||| | ||||||||||||||  |||
Db   5198 TCGCGGCTATGTCGCCGAGGCGACCGGCGCCAACGTGTTCTTCGTCAAGGACGGCGTGCT 5139

Qy    732 GGCCTCCCCGTCCCGGAACGCGTTGCCG---GGCATCACCCGTAAGACGGTGTTCGAACT 788
          |  | | ||    |  ||   || | |   ||||||||||    |||||| |||| ||
Db   5138 GCACACGCCCAAGCCCGACTGCTTCCTGGACGGCATCACCCGCCGCACGGTGATCGACCT 5079

Qy    789 GGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGA 848
          |||| |   | ||| |||||  |    |  |||| || ||   |     |||||   |
Db   5078 GGCCAAGGCCAAGGGGATCGAGGTGGTCGAGCGCCACATCCAGAAGGAAGAACTGGCCAC 5019

Qy    849 CGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATC 897
          |  |  |||||   |    |||  |||||| | || ||| || ||  ||
Db   5018 CTTCACCGAGTGCTTCATCGTCGGCACCGCCGCCGAGGTGACCCCAGTC 4970


RESULT 6
AE010451/c
LOCUS       AE010451               10685 bp    DNA     linear   BCT 01-AUG-2002
DEFINITION  Methanopyrus kandleri AV19 section 150 of 157 of the complete
            genome.
ACCESSION   AE010451 AE009439
VERSION     AE010451.1  GI:19888323
KEYWORDS    .
SOURCE      Methanopyrus kandleri AV19.
  ORGANISM  Methanopyrus kandleri AV19
            Archaea; Euryarchaeota; Methanopyri; Methanopyrales;
            Methanopyraceae; Methanopyrus.
```

REFERENCE 1 (bases 1 to 10685)
AUTHORS Slesarev,A.I., Mezhevaya,K.V., Makarova,K.S., Polushin,N.N., Shcherbinina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L., Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O., Malykh,A.G., Koonin,E.V. and Kozyavkin,S.A.
TITLE The Complete Genome of the Hyperthermophile Methanopyrus kandleri AV19 and Monophyly of Archaeal Methanogens
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10685)
AUTHORS Slesarev,A.I., Mezhevaya,K.V., Makarova,K.S., Polushin,N.N., Shcherbinina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L., Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O., Malykh,A.G., Koonin,E.V. and Kozyavkin,S.A.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2002) Fidelity Systems, Inc., Gaithersburg, MD 20879

FEATURES Location/Qualifiers

source 1. .10685
/organism="Methanopyrus kandleri AV19"
/strain="AV19"
/db_xref="taxon:190192"

gene complement(123. .1466)
/gene="ffh"
/note="MK1612"

CDS complement(123. .1466)
/gene="ffh"
/codon_start=1
/transl_table=11
/product="Signal recognition particle GTPase"
/protein_id="AAM02825.1"
/db_xref="GI:19888324"
/translation="MIGFADRLAEITKKIKGASIIDEDFVKEVVRDVQRALLEADVDV
KLVLELSKRIEKRALEEEPPAGVPKRDYLLRIVYEELVELLGGEKTEGLDIDLSRDVN
VIMLVGLYGMGKTTTAAKLARYLQRKGYRVGLVGADPYRPAAGEQLRQLAEEVDVPVH
VEDVDDAVEMAVKGVEALKDECDVVIVDTAGRDRLSEDLIDELREMAERIEPHEVLLV
LDATVGQKAGDHAEAFHEAVQLTGVVITKLDTAAKGGGALSAVARTGAPIKFVGTGER
VDDLEEFNPRSFVARLLGIGDIDELLRRTEEMLEEEEKAEDVLEGEFTLKDLYEQLEA
LSKMGPVDKLLQYVPGMGGGRNVRKISQITEERLKKYKVIMDSMTEKELENPEILNKS
RIRRIAIGSGTSERDVIELLNHYRMMKDVIEDIQSGRIPRIGGELGRVIRNVLRG"

gene complement(1463. .2512)
/gene="ftsY"
/note="MK1613"

CDS complement(1463. .2512)
/gene="ftsY"
/codon_start=1
/transl_table=11
/product="Signal recognition particle GTPase"
/protein_id="AAM02826.1"
/db_xref="GI:19888325"
/translation="MPSSLFGKMKSVLSKVKEKVAKRVEERAEEAEEAKVKEVETMPE
DLKKPTLKERLKRVVKREVTITEADIEDILDELELELLISNDVAVEVAESIREELKKEL
VGRRVKGKSEIPKVVEEGFREALLSVLEPKKEVDLMETVEKARQDGRPAIIMFVGVNG
SGKTTTIAKVAKLLKDHGYSVVIAAADTFRAAAIEQLEEHAERLGVTLIKGERGDDPT
AVAFNAVQHAEAKGKDVVLVDTAGRAYTDVNLMEELKKMKRVLEPDLVVFVGDALAGN
DAIEQAKTFHEYVGIDCAILTKVDADAKGGAVLSISKVTGAPILYLGVGQDYDDLKAF
SPEWFVERVIGGEES"

gene complement(2542. .3015)
/gene="gim5"
/note="MK1614"

CDS complement(2542. .3015)
/gene="gim5"
/codon_start=1
/transl_table=11
/product="Predicted prefoldin, molecular chaperone implicated in de novo protein folding"
/protein_id="AAM02827.1"
/db_xref="GI:19888326"
/translation="MAEKKNEQEIQQELQRLIAEINRLQGQMEAINAQIDLIESSISE
LNRVEETLKGVKELEGDEEVLVPVGAQSFVRACVTDTERVIVGIGAGVAVERTIDEAL
ESIDDQRQELEKARAEAQQKLQELAQELQEKQRKAQELAQQLEGAQRIAQQSGGG"

gene complement(3015. .3257)
/gene="rpl20A"
/note="MK1615"

CDS complement(3015. .3257)
/gene="rpl20A"
/codon_start=1
/transl_table=11
/product="Ribosomal protein L20A (L18A)"
/protein_id="AAM02828.1"
/db_xref="GI:19888327"
/translation="MSEVKVFEVRGTFRMGDEPRQPFTRQVPATSEEEALEKVYSDLG
SEHGVSRMEIQIEEIREIDPSKVEDPILRRLLGVEE"

gene complement(3275. .3940)
/gene="tif6"
/note="MK1616"

CDS complement(3275. .3940)
/gene="tif6"
/codon_start=1
/transl_table=11
/product="Translation initiation factor 6 (EIF6)"
/protein_id="AAM02829.1"
/db_xref="GI:19888328"
/translation="MTVVKASVHGDPNIGAWIAASEEYAVVAPKVPDDIVERVKEALD
VEVVRTTVAGSNLVGALLAVNSNGALFPRHAREHEIRVVRELGVEVDVLPSKMNAVGN
LVLTNDHGALVHPDLDDHALEVIESVLGGRVVRGELGGVKTVGSAGVANSKGAVVHPG
ATEEEMERVSEVLGVDVEVGTVNRGSPYVGVGIVVNSKGAVVGEDTTGPELARLEDAL
YLI"

gene complement(3943. .4203)
/gene="rpl31A"
/note="MK1617"

CDS complement(3943. .4203)
/gene="rpl31A"
/codon_start=1
/transl_table=11
/product="Ribosomal protein L31E"
/protein_id="AAM02830.1"
/db_xref="GI:19888329"
/translation="MAEVVDERVYTVPLRDAKKAPLKKRAPRAVKALRQFIERHMKAE
EVRIGNDVNEKIWERGIKKPPSKIRVRAVKYADGTVEVRLAE"

gene complement(4219. .4374)
/gene="rpl39"
/note="MK1618"

CDS complement(4219. .4374)
/gene="rpl39"
/codon_start=1
/transl_table=11
/product="Ribosomal protein L39E"
/protein_id="AAM02831.1"
/db_xref="GI:19888330"
/translation="MARVKPLGKKLRMAKAIKQNRRVPPWVVAKTGGRVIDNPKRRHW
RRSKLKP"

gene complement(4388. .4741)
/gene="MK1619"

CDS complement(4388. .4741)
/gene="MK1619"
/codon_start=1
/transl_table=11
/product="DNA-binding protein"
/protein_id="AAM02832.1"
/db_xref="GI:19888331"
/translation="MTDPELERIRRKKIMELQRKLEESQEKKVEEEREKKALEEAQRR
AMLRRILTPEARERLARVRLARPQLAQAVENYLLQLAQTGQLKEKIDEDQLKRILKQV
SDATRKEYRIRFKRK"

gene complement(4771. .5220)
/gene="rps19A"
/note="MK1620"

CDS complement(4771. .5220)
/gene="rps19A"
/codon_start=1
/transl_table=11
/product="Ribosomal protein S19E (S16A)"
/protein_id="AAM02833.1"
/db_xref="GI:19888332"
/translation="MYDAYVVPGSELVERLAEKLKDFEEIKPPEWAKYVKTGRHKERP
PEDPDWWYMRAASILRRVYMDGPVGVSRLRTYYGGRQDRGARPERFRKGSGAIIRKIL
QQLEEAGLVEKTEEGRVVTPEGRSLVDSTAHEIAKEKGYTDKFTSPI"

```
     gene            complement(5266. .5583)
                     /gene="MK1621"
     CDS             complement(5266. .5583)
                     /gene="MK1621"
                     /codon_start=1
                     /transl_table=11
                     /product="Predicted RNA-binding protein containing KH
                     domain, possibly ribosomal protein"
                     /protein_id="AAM02834.1"
                     /db_xref="GI:19888333"
                     /translation="MTQVKRLSGKERRALRARAVLLDPVVRIGKKGLTSGVIQEVDRQ
                     LEERGLIKVRFERNILRRYDRKELAEELARKVNAELIDVRGRTAVLFRPREGWRRFHG
                     LSR"
     gene            complement(5570. .5956)
                     /gene="rpr2"
                     /note="MK1622"
     CDS             complement(5570. .5956)
                     /gene="rpr2"
                     /codon_start=1
                     /transl_table=11
                     /product="RNAse P subunit RPR2"
                     /protein_id="AAM02835.1"
                     /db_xref="GI:19888334"
                     /translation="MCKVESPRSGVLLLRRIALERAERLLRLARTVYYEDPDRARRYV
                     ELARRIAMKARVKLPKHLKRSFCKRCNTPLIPGVTARVRLRQNRMPHVSVTCLECGYI
                     YRYPYLREVKERRRRHMEGVKDRDAG"
     gene            5910. .7145
                     /gene="MK1623"
     CDS             5910. .7145
                     /gene="MK1623"
                     /codon_start=1
                     /transl_table=11
                     /product="Uncharacterized protein specific for M.kandleri,
                     MK-39 family"
                     /protein_id="AAM02836.1"
                     /db_xref="GI:19888335"
                     /translation="MRNSNTPERGDSTLHTWLLLWLLLPMTPAQADPTSHYLEVHPDT
                     VPVTPEGSPLVPAYVELTDEGKEAALSLLERLGIPSDEAAVGYLADVFVGESCHGGTR
                     CLVVYNPQDTTVEVRPWEEDRLPLRLPPLRLDDHLTELLLCLPVREQITLQPAVLELG
                     WNLEMMQRCTKAVRLHKFGEPTAVIAVYVPTETAALRAFESLLPDPRNVARAMLERLS
                     SEVRDAIDEYLVRDRLERLVKSRLGSDVYDRVREYFSSMIHLTVSKIEELYDERVLPA
                     LIEWCAAIVRYVETSLITAFALLYPDNPSADVLERRVAVYNSPYFQILRDWLLEPSYT
                     AWKSVLCDLAVTVADSIFEAARPVLSALRLDSESSRTTLEEITKSAVKNFIEALGPCL
                     RTVLELLLLDTLMGLAWAC"

  Query Match             12.4%;  Score 126.6;  DB 1;  Length 10685;
  Best Local Similarity   48.4%;  Pred. No. 1.2e-08;
  Matches  419;  Conservative    0;  Mismatches  434;  Indels   12;  Gaps    2;

Qy    140 TCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACA 199
          ||  ||   |  | |||||||    | ||    ||||| || || ||||||| ||| ||||
Db  10546 TCATCTACCTGAACGGCGAACTGGTCCCCCGGGAAGAGGCTAAAATCTCCGTCTACGACC 10487

Qy    200 CCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCT 259
          ||| |||    |   |||      || |   |     |  |  |   | |||   || |
Db  10486 ACGGGTTCCTGTACGGCGACGGCGTCTTCGAGGGGATTCGGGCCTACGATGGCCGGATAT 10427

Qy    260 TCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACG 319
          ||  | | | ||| ||| | ||| |  ||  ||||   ||         |    || ||
Db  10426 TCAAGTTGGACGAGCACGTCGACAGACTGTACGACTCCGCAAAGGCCATCATGCTCGAGA 10367

Qy    320 CCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGC 379
            |   | ||||||  |  ||   |||  || ||   ||  |||   |        |||
Db  10366 TACCGATGACCAAGGAGAAGATGAAGGAGGCCATCATCGAGACCGTACGGGCCAACGAGC 10307

Qy    380 TGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGA 439
          | || ||  ||| | |     ||   ||| | || ||  | ||  | | |   |    ||
Db  10306 TCCGGGACGCGTACATCCGGGTGGTGGTCTCGCGCGGCGAAGGTGACCTCGGGTTGGACC 10247

Qy    440 AGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCT 499
           |||   ||    |   || |  ||||          |||| || | | |  ||
Db  10246 CGGAGAAGTGTCCGGAACCCAACGTGGTG---------ATCATCGCGGAGCCGATGGAGC 10196

Qy    500 TCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCG 559
            |  ||  || | |  |          | || | |  |  ||  |   || || ||
Db  10195 CGTTGTACGGGGACCTGTACGAGAAAGGGATCGAGGTGATCACGGCCTCCGTGCGTCGTA 10136

Qy    560 CCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAG 619
            |   || |  |||| ||||  ||||||   | | | |       |   || |   |
Db  10135 TTCCACCGGACGCGCTCGATCCGAAGATCAAGTCGTGCAACTACCTGAACAACATCCTGG 10076

Qy    620 CCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACT 679
          | |   || ||||||||       || ||   |   || ||| || | |||   ||   ||
Db  10075 CGAAGATCCAAGCCAACCTGGCGGGCGCCGACGAGGCCATCATGTTGGACCACGAGGGCT 10016

Qy    680 GCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCC 739
           |||   ||| ||| | ||   |||||||| | | ||  | |||||  |   ||  | | |
Db  10015 ACGTCTGCGAGGGTACCGGTGACAACGTGTTCGTGGTGGAAGACGGGACCGTGTACACTC 9956

Qy    740 CGTCCCGGAACGCGTTGCC---GGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACC 796
          |  |   | || |  | |    ||| |||||| |   ||| ||| | ||  |   |||
Db  9955 CACCGGAGGACACTATTCTCCGGGGGATCACCAGGGCGACCGTGATGGAGATCTGCGAGG 9896

Qy    797 AGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACG 856
          || | || ||   |         |       | ||     | ||||| |||| || ||||
Db  9895 AGCTCGGTATACCCGTGGAAGAGAAGCGGATAACACTGGGAGAACTGTACGCGGCGGACG 9836

Qy    857 AGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGG 916
          || || |          ||| ||||  | |||  | ||  | |    | | || ||   |
Db  9835 AGGTGTTCCTAACGGGCACGGCGGCGGAGGTGGCCCCCGTTAGGAAGGTCGACGGGCGGA 9776

Qy    917 CCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGC 976
            | ||| | |     || || ||  | |||  |   |||  |||   |||| ||| ||
Db  9775 AGATCGGCGAGGAGTGCCCAGGACCCATCACGAGGCGTATCATGGAGGCGTTCCGGGAGC 9716

Qy    977 TGATGGACGAGCCGGGCCCGCTGAT 1001
          | | | |  ||   ||  ||| | |
Db  9715 TCACGAAGAAGGAAGGTACGCCGGT 9691
```

```
RESULT 7
AL646080
LOCUS       AL646080              190050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 5/11.
ACCESSION   AL646080 AL646053
VERSION     AL646080.1  GI:17431113
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 190050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
  COMMENT   Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
```

```
FEATURES             Location/Qualifiers
     source          1. .190050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
                     /plasmid="megaplasmid"
     gene            51. .362
                     /gene="RSp0643"
                     /note="RS05579"
     CDS             51. .362
                     /gene="RSp0643"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAD17794.1"
                     /db_xref="GI:17431114"
                     /translation="MKTSPSPNPAAAGPASDPEAAGDGQVAFEQLHEYRNHTITIRPR
                     RIGGGLWRAAFAVSLDSRALVAVGTEVGQAFRTPIGAARDAVDAARSFIDARLKACRP
                     A"
     gene            397. .555
                     /gene="RSp0644"
                     /note="RS05578"
     CDS             397. .555
                     /gene="RSp0644"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
                     /protein_id="CAD17795.1"
                     /db_xref="GI:17431115"
                     /translation="MERQQLFAIAIVIAAVSAAYSLARDLNDDERDFPPGQAALVHGH
                     ATASAAGR"
     gene            complement(575. .1135)
                     /gene="RSp0645"
                     /note="RS05577"
     CDS             complement(575. .1135)
                     /gene="RSp0645"
                     /function="miscellaneous; not classified regulator"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"
                     /protein_id="CAD17796.1"
                     /db_xref="GI:17431116"
                     /translation="MKRISTFPDFVDMKKVAIKFRSPAAAPPQGAALSALDRTDRAIL
                     KWLQRDASISNVALADKVSLSPPACLRRVERLKELGCIRGVVALLEPKALDLGTLVMI
                     GVVLDRSTPESFADFEKAVQKVPGCLECHVVTGEFDYFMLVRTRDNDSYNRLHAEQLL
                     YLPGVRQIRTFMTLKQVLSTTQLPIA"
     gene            1244. .2260
                     /gene="RSp0646"
                     /note="RS05576"
     CDS             1244. .2260
                     /gene="RSp0646"
                     /EC_number="4.1.99.4"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE
                     DEAMINASE (ACC DEAMINASE) PROTEIN"
                     /protein_id="CAD17797.1"
                     /db_xref="GI:17431117"
                     /translation="MNLNKHPRHPLTFGPTPIQPLKRLSAHLGGKVELYAKREDCNSG
                     LAFGGNKTRKLEYLVPEVLAGGYDTLVSIGGIQSNQTRQVAAVAAHLGLKCVLVQENW
                     VNYADAVYDRVGNIELSRILGADVRLDAAGFDIGIRPSWEQAMEDVRRAGGKPFPIPA
                     GCSEHPLGGLGFVGFAEEVRQQEAEFGFRFDYIVVCSVTGSTQAGMVVGFAADGRADR
                     VIGIDASAKPEQTREQILRIARDTAKLVELGRDITEDDVVLDTRYGGPEYGLPNEGTL
                     EAIRLCARQEGMLTDPVYEGKSMHGMIDRVRGGEFPEGSRVLYAHLGGVPALNAYSFL
                     FRNG"
     gene            complement(2336. .3508)
                     /gene="RSp0647"
                     /note="RS05575"
     CDS             complement(2336. .3508)
                     /gene="RSp0647"
                     /EC_number="4.2.1.17"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"
                     /protein_id="CAD17798.1"
                     /db_xref="GI:17431118"
                     /translation="MSALPKPAEPAAAAEPEVLFHVVNRVAILTLNRPSALNALSHGM
                     VRQLSALLAQCRDDPRIAAVVLRGAGEKGFCAGGDVRALYRLARQGQAHGPDGWLQFF
                     VDEYRLDYALHRFPKPVVALLDGIAMGGGMGLGQGARLRVVTERSRIAMPETRIGFLP
                     DVGATHFLGAMPPEMALYIGLTGTTLSGADALHCGLADVCVPAEWLASFEDRLLRLSH
                     AGDPGDLMQALRRVFEPPCNIVPHAAIARTTPWIVRYFDRRSTIDRIVATLRESLERD
                     PAREARQWLQATLDAMATHSPTMLHVTREALLRGRQLTLAECFRMELGIVARAIEEGD
                     FCEGVRAHLVDKDHRPGWAPATLAGVRPERVRHFLSSPWRGAAHPLAHLGAQPQAA"
     gene            complement(3505. .4305)
                     /gene="RSp0648"
                     /note="RS05574"
     CDS             complement(3505. .4305)
                     /gene="RSp0648"
                     /EC_number="4.2.1.17"
                     /function="small molecule metabolism; degradation; carbon
                     compounds"
                     /note="Product confidence : probable
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE ENOYL-COA HYDRATASE PROTEIN"
                     /protein_id="CAD17799.1"
                     /db_xref="GI:17431119"
                     /translation="MIALDWLHDGTVALLTLSRPPANAFTPEGLRQLQATVETLDADP
                     RVRALVITGDGPKFFSAGADLNTFASGDRDVAREAAARFGAAFEALQNARPVVIAAIN
                     GYAMGGGLECALACDIRIAERHAQLAVPETAVGLLCCGCGTQTLPWLVGEGWAKRMIL
                     TGERIDAQTALRIGLVEEVVDTGAAREAALAMARRVTTLSPRAVTFSKALIHQARHGV
                     PRTAALAVERERFVDLFGHPDQREGVNAFLEKRPPRWHTAARPEEALP"
     gene            complement(4302. .5198)
                     /gene="mmsB"
                     /note="RSp0649; RS05573"
     CDS             complement(4302. .5198)
                     /gene="mmsB"
                     /EC_number="1.1.1.31"
                     /function="small molecule metabolism; amino acid
```

```
                     biosynthesis; isoleucine/valine"
                     /note="Product confidence : probable
                     Gene name confidence : putative
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE 3-HYDROXYISOBUTYRATE DEHYDROGENASE
                     OXIDOREDUCTASE PROTEIN"
                     /protein_id="CAD17800.1"
                     /db_xref="GI:17431120"
                     /translation="MKIAFIGLGNMGAPMARNLLKAGHALSVFDLDARAVQSLVDAGA
                     SAAPSPKAAATGADCAITMLPAAAHVHSVLTADDGILAGIAKGVPIIDSSTIDPDSAK
                     ALAALATAHGNPFVDAPVSGGTGGAAAGTLTFMVGGGAAVFEQVRPVLSAMGKNLVHC
                     GEAGAGQGAKICNNLVLGITMAGVAEAMSLGEALGIDPKVLGGIINTSTGRCWSSDTY
                     NPFPGVIETAPSSRGYTGGFGTDLMLKDLGLAGDAAKSVRQPVYLGALAQQLYQASSS
                     QGAGKLDFSAVIRLYRPAGAAR"
     gene            complement(5220. .6767)

  Query Match             10.8%;  Score 109.8;  DB 1;  Length 190050;
  Best Local Similarity   47.7%;  Pred. No. 1.2e-06;
  Matches  398;  Conservative    0;  Mismatches  422;  Indels   15;  Gaps    2;

Qy     135 CGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTT 194
           | |  ||| ||   |  | ||| ||||| |||||||||  || || |   |||||||| |
Db   82040 CCGAATCGTCTTCCTGAACGGCCAATACGTGCCGGCCGCCGAGGCCACCGTCTCCATCAT 82099

Qy     195 CGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAA 254
           |||  ||| |||     | |||||   | ||||   ||  |    |||   | ||||||
Db   82100 GGACCGCGGCTTCACCTTTGCCGATGGCATCTACGAAGTGACCGCGGTAGCACGCGGCAA 82159

Qy     255 CATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCT 314
             |  ||     || || |||||| |  ||  ||     |  |  |||| || |  || |
Db   82160 GCTGGTCGACAACGACGCCCACCTCGCGCGCCTGACGCGCTCGCTGTCCGAGATCGGCAT 82219

Qy     315 GGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTC 374
           |||  |  ||||| | | | ||||  | |     |   | ||  |    |  ||
Db   82220 CGACAACCCGTACACCGATGCCGAATGGACGCGCGTGTGCGAGGCGCTGATCGCGCGCAA 82279

Qy     375 GCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGG 434
             |||   ||     | ||| |  ||   ||||| || ||     |  |||| |
Db   82280 CGGCCTGGAAGAGGGCGTGGTGTACATGCAGGTCACGCGCGGCGCGGCCGAGCGCGACTT 82339

Qy     435 CGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTG 494
           ||  |    |  |   |      ||     ||||    ||        ||  || | |
Db   82340 CGGCATCCCCGCGCAGATCGCGCCCACGGCGGTGGCGTTCACGCAGGTCAAGTCCATCGT 82399

Qy     495 GGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCG 554
           | |  |||||  ||      |  |   ||| |  | | | |   |  || | || | ||
Db   82400 CGACAGCCCGCTGGC------GCGCAAGGGCGCGGCGGTGGTGACGGTGCCCGATCTGCG 82453

Qy     555 CCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCAC 614
           | |     || || ||| |     |       |   |     |      ||   | |
Db   82454 CTGGAAGCGCTGCGACATCAAGAGCGTCGGGCTGTTGCCGCAGGTCATGGCCAAGCAGGT 82513

Qy     615 CGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGA 674
           | ||| |   ||  ||||  ||     |   |  |  ||          ||||   |
Db   82514 GGCCGCCCGCGCCGGCGCCAGCGAAGCCTGGATGACCGACG---------GCGACCGCGT 82564

Qy     675 CAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGC 734
           | |   |  |||      | | ||||| | |   || || ||||  |   |     |
Db   82565 GACCGAGGGCGCCTCGTCCACCGCCTTCATCATCACCACCGACAAGCGCCTGATCACGCG 82624

Qy     735 CTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGA 794
             | | |||       | | |||| |||||||| ||      ||||| | |  |||||
Db   82625 TCCGCTGTCGAATGCGGTGCTGCCCGGCATCACGCGCGTCTCGGTGATGGCGCTGGCACG 82684

Qy     795 CCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGA 854
           | ||   ||| | |          ||||  | |||||   |  |||    |  | || |
Db   82685 CGAGCACGGCCTGGTGCTGGAGGAGCGCACCTTCACCGTGCAGGAAGCGCAGCAGGCGGC 82744

Qy     855 CGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGA 914
           ||||  | |      || | |  |   |   || |  ||  ||| |||  | |||||||
Db   82745 CGAGGCGTTCTTCACCAGCGCTTCCACCTTCGTGATGCCCGTCATCTCCATCGATGGCGT 82804

Qy     915 GGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTC 969
             | |||||| ||||| |||| || || || |||   ||  | ||    | || |
Db   82805 ACCGGTGGGCGACGGCCAGCCGGGCCCGCTCACGCGCGCACTGCGCACGCTGTAC 82859


RESULT 8
AE005101/c
LOCUS       AE005101               12730 bp    DNA     linear   BCT 12-FEB-2001
DEFINITION  Halobacterium sp. NRC-1 section 132 of 170 of the complete genome.
ACCESSION   AE005101 AE004437
VERSION     AE005101.1  GI:10581532
KEYWORDS    .
SOURCE      Halobacterium sp. NRC-1.
  ORGANISM  Halobacterium sp. NRC-1
            Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE   1  (bases 1 to 12730)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Genome sequence of Halobacterium species NRC-1
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
  MEDLINE   20504483
   PUBMED   11016950
REFERENCE   2  (bases 1 to 12730)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES             Location/Qualifiers
     source          1. .12730
                     /organism="Halobacterium sp. NRC-1"
                     /strain="NRC-1"
                     /db_xref="taxon:64091"
     gene            complement(70. .597)
                     /gene="pyrE2"
                     /note="VNG2118G"
     CDS             complement(70. .597)
                     /gene="pyrE2"
                     /note="PyrE2"
                     /codon_start=1
                     /transl_table=11
                     /product="orotate phosphoribosyl transferase"
                     /protein_id="AAG20258.1"
                     /db_xref="GI:10581533"
                     /translation="MSATDDLVSALRAADAVQFGEFELSHGGTSEYYVDKYLFETDPE
                     CLSAIAAAFADRIDEDTTLAGVALGGVPLAAATATEAGVPYVIARKQAKEYGTANRIE
                     GRLDDGEEVVVVEDIATTGQSAVDAVDALRDAGATVNRALIVVDREEGGRELLAEHGV
                     EMAALVTASDLLDAE"
     gene            complement(653. .1297)
                     /gene="VNG2119C"
     CDS             complement(653. .1297)
                     /gene="VNG2119C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
```

```
                /transl_table=11
                /product="Vng2119c"
                /protein_id="AAG20259.1"
                /db_xref="GI:10581534"
                /translation="MGGIIPTSRSPHVLNTRPLGIRGFLPGARPPTRMDLVGTVVVAV
                WAMLPAYVPNNAAVLAGGGRPIDGGRSLGGRRLLGDGKTWRGTAVGTAAGVALAVALN
                ALRPAAADALGVVLPAFPPAAMGTLAFGAMVGDIAASFLKRRTGRQRGAAFPVVDQLD
                FVVVALALTALAVPAWVGDTFGLPVLVTVAVLTPALHLLTNGIAYALGVKDEPW"
gene            1284. .2123
                /gene="yusM"
                /note="VNG2120G"
CDS             1284. .2123
                /gene="yusM"
                /note="YusM"
                /codon_start=1
                /transl_table=11
                /product="proline dehydrogenase"
                /protein_id="AAG20260.1"
                /db_xref="GI:10581535"
                /translation="MIPPIANRFVAGETPAEAIEYAADLNDRGVTAILNLLGEHYHER
                PPADEDAAAYRRLVADVGDSDVDACVSVKPSQIGLDVGPEVFAENLETIAAAAADHDV
                FAWVDMEDHETTDATMDTFESLARTHAGNMGLCVQANLKRTEQDLDRLADVPGKVRLV
                KGAYDEPDDVAYQDKPDVNEAYRTHLEFMFREFDDGVAVGSHDPAMISYAAELHEEYG
                TDYEVQMLMGIREDAQTELAATGVPTYQYVPYGGKWFSYFYRRAMERKENLVFALRAV
                LGR"
gene            2195. .2812
                /gene="VNG2121C"
CDS             2195. .2812
                /gene="VNG2121C"
                /note="conserved hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng2121c"
                /protein_id="AAG20261.1"
                /db_xref="GI:10581536"
                /translation="MGSWKRDFASGLIVLVPIVVTLYVIYWLFGLLSNITLFTQLTDT
                QYQAVAATLAVFVLIVFSVGYLMRTAAGSILEALIDYVMNRLPVLRIVYNASKMAVET
                VLSDGTGEFQQPVRVEPWPGMRLTAFKTGKKAPDGREVVFMPTAPNITTGFVMEVEPE
                DLEETDEGVEDALTRVLSAGFGENDDSEVAIDELLPEDDDDENEE"
gene            complement(2850. .3776)
                /gene="ilvE2"
                /note="VNG2122G"
CDS             complement(2850. .3776)
                /gene="ilvE2"
                /note="IlvE2"
                /codon_start=1
                /transl_table=11
                /product="branched-chain amino acid aminotransferase"
                /protein_id="AAG20262.1"
                /db_xref="GI:10581537"
                /translation="MPSFSEMDVDTIWMDGEFVDWEDAQVHVLTHALHYGSGIFEGVR
                VYDTDNGPAIFRWDDHLQRFYESAKPYDLDIEFTPDELTDATVELIQRQDLGSCYIRP
                LAYYGYESLGVSPGDCPTDVAIAAWPWGAYLGDDALENGVDVMVSTWRKHASSQIPTN
                AKTTGLYVNSMLAGEEARRNGFTEAIVLNKEGNVAEGPGENIFLVRDGELFTPGLSES
                ILDGITRDTVITLAEERGYTVHDDVSISRGELHTADELFFTGSAAEVTPIKQVDNVEI
                GSGTRGPVTEELQTAFFDLVEAGDREEWFHYV"
gene            3908. .5383
                /gene="nhaC2"
                /note="VNG2123G"
CDS             3908. .5383
                /gene="nhaC2"
                /note="NhaC2"
                /codon_start=1
                /transl_table=11
                /product="Na+/H+ antiporter"
                /protein_id="AAG20263.1"
                /db_xref="GI:10581538"
                /translation="MSDFTPLAYEDIDDARKPTLTEALVPVLVVVVSLGVGSGYLGLA
                PHAPLLWSITFTGLFARYHLGYDWERVSDAALAGLRMGLQAILILFVIYGLIATWTSA
                GTIPGLMYYGLGALTPAVFLPVTAILAAVVAFAIGSSWTTVGTLGVAFVGIGSGLGIP
                LPMTAGAIVSGAYAGDKQSPLSDTTNLAAAVTNADLYDHINGMRLGTAIAFGLSVLAY
                AVLGVLIVDGGAANVTAISGPLAGSYALGPLVFLPLVVTFGLAFRGYPALPSLVAGVF
                AGALTTIAVQGASFTAAWDIFLNGTTPETGSELVNNLLTTGGIAGSAWTIAVVVAALS
                LGGLLERTGVLPTLTHHLTRAIWSPESLVAGTGVAAIATNAFSAQQYMSIVVPSVSLR
                TLYDEYDLNERDLSRAVEAAGTPTGPFFPWHAGAVFMTGALYTSQPGAISWWWAPYYF
                FGILSPLVLFGLVLAGRGYDRAGSATQPAAPAANAAATDDD"
gene            complement(5385. .7133)
                /gene="VNG2124C"
CDS             complement(5385. .7133)
                /gene="VNG2124C"
                /note="conserved hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng2124c"
                /protein_id="AAG20264.1"
                /db_xref="GI:10581539"
                /translation="MASYGVLALVPPLLAIVLAMTTRQVLVSLFAGVWAGALVVAQWN
                PIAATALTMDWLVAVVRTPFNTKFIVLILFMGAGAAFIYRSGGILALERWIGSRISTA
                RESQILTWLIGVFIFFDSYTSTVVTGNATRELAQDADSSREMHAYVLDSTTAPVTTFG
                PVSNWIGYQVSMIIAGLEAARVSAAELGTTPFGLFLQSIPWNIYCFMAFFMVGFITIT
                QRFFGPMLDAEWRARTTGDTIADGATPLSDVSNDVGDPSTANPSLLNFFGPIAVLLAV
                GLGSMWYLGGGYKPGVDIATALQETDVAMGLLHGAFAFMVSGFVGALAYGTMDLEGAT
                DTIVDGFKTMNIALAIIVLAWGIGMAAEKVGTADFIVTTLVGSGVPAGFLPLLVFLAA
                MFVAFTTGTSWGTMAILTPVAIPLGYNLAGAAVLPVLVGVLFGGAIWGDHSSPISDTT
                VMSSIFAGSDHIDHVTTQIPFAATAAGVTIVMLVLYAVGLRSPVVALPLATVLTLGAV
                VVLNKLDTRRKGLPEVMPRADAVDEADREAAANGTLSLPAAATSYDYLSLVPVAAVGI
                VLLYLSLVFVFAALGG"
gene            complement(7210. .7716)
                /gene="ribA"
                /note="VNG2125G"
CDS             complement(7210. .7716)
                /gene="ribA"
                /note="RibA"
                /codon_start=1
                /transl_table=11
                /product="GTP cyclohydrolase II"
                /protein_id="AAG20265.1"
                /db_xref="GI:10581540"
                /translation="MRTDAGGLICVALGDAVADAFALPFIADELDHPAADHGHLGYDD
                RSSFSLTVNHRDTYTGITDDDDRALTIAELGQAAADPAAVDFPATFRAPGHVHVLRAAP
                SLSARQGHTELGIALAGAAGVPPAVVVCEMLDADTGGARSVADAKRYASRHDIPYVEG
                SDLVAAFQ"
gene            complement(7865. .8593)
                /gene="VNG2126C"
CDS             complement(7865. .8593)
                /gene="VNG2126C"
                /note="conserved hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng2126c"
                /protein_id="AAG20266.1"
                /db_xref="GI:10581541"
                /translation="MRGAGQPPAMSGATSTGDVGYDELAVLKLLALDGAHRGEVKVSC
                GDLASRLDASSQTASRRLQALDDADHVTRDLVSDGQWITVTDAGRHALKHEYEDYRRI
                FEDPGELALAGTVTSGMGEGRHYISLPGYNRQFAEKLGYEPYPGTLNVDLPPDGQRAR
                AGIQALDGVDIDAWEDEDRTYGSATCYACTVVGDGTDFDGAHVIIPDRTHHDDDQLEI
                IAPVKLRERLGLLDDDEVTIRVEA"
gene            complement(8630. .9607)
                /gene="VNG2128C"
CDS             complement(8630. .9607)
                /gene="VNG2128C"
                /note="conserved hypothetical protein"
                /codon_start=1
                /transl_table=11
                /product="Vng2128c"
                /protein_id="AAG20267.1"
                /db_xref="GI:10581542"
                /translation="MSESDGPKQVDDPDYHHENHTAAQTCGWTANAMRGEGTCYKHAF
```

```
Query Match          9.4%;  Score 95.8;  DB 1;  Length 12730;
Best Local Similarity  47.7%;  Pred. No. 0.00019;
Matches  428;  Conservative    0;  Mismatches  442;  Indels   27;  Gaps    4;

Qy    110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169
          ||  ||  | |||   | | | || || |  |  ||||| || ||||| | | |
Db   3775 TGCCCAGTTTCAGTGAGATGGACGTCGACACCATCTGGATGGACGGCGAGTTCGTCGACT 3716
```

```
Qy    170 CCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACA 229
Db   3715 GGGAGGACGCACAAGTACACGTCCTCACCCACGCACTCCACTACGGGTCCGGCATCTTCG 3656

Qy    230 CCGTCGCGCATGTATGGCACGGCAAC----------ATCTTCCGGCTCGGCGACCACC 277
Db   3655 AGGGGGTCCGTGTCTACGACACCGACAACGGCCCCGCGATCTTCCGGTGGGACGACCACC 3596

Qy    278 TGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACG 337
Db   3595 TCCAGCGGTTCTACGAATCGGCGAAACCCTACGACCTCGACATCGAGTTCACGCCCGACG 3536

Qy    338 AACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGA 397
Db   3535 AACTCACCGACGCCACCGTCGAGCTCATCCAGCGCCAAGACCTCGGGTCCTGTTACATCC 3476

Qy    398 ATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCA 457
Db   3475 GCCCGCTGGCGTACTACGGCTACGAATCACTCGGCGTCTC------CCCCGGCGACTGCC 3422

Qy    458 CCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGA 517
Db   3421 CGACCGACGTCGCCATCGCCGCCTGGCCGTGGGGCGCGTACCTCGGCGACGACGCCCTCG 3362

Qy    518 TCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCG 577
Db   3361 AAAACGGCGTCGACGTGATGGTGTCGACGTGGCGCAAGCACGCCTCCAGCCAGATCCCGA 3302

Qy    578 ACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGG 637
Db   3301 CGAACGCCAAAACCACCGGCCTCTACGTCAACAGCATGCTCGCCGGCGAGGAAGCCCGGC 3242

Qy    638 ACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGG 697
Db   3241 GCAACGGCTTCACCGAAGCCATCGTCCTCAACAAAGAGGGCAACGTCGCGGAAGGCCCCG 3182

Qy    698 GCTTCAACGTGTGCATCGTCAAGGACGGCAAGCT------GGCCTCCCCGTCCCGGAACG 751
Db   3181 GCGAGAACATCTTCCTCGTGCGCGACGGCGAACTCTTCACGCCCGGCCTCTCCGAGAGCA 3122

Qy    752 CGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAG 811
Db   3121 TCCTCGACGGCATCACCCGCGACACCGTCATCACGCTCGCCGAGGAGCGCGGCTACACCG 3062

Qy    812 CCACCCTGCGCGACGTCACCAGCCGT---GAACTCTACGACGCCGACGAGTTGATGGCGG 868
Db   3061 TCCACGACGACGTCTCGATCTCGCGTGGCGAACTCCACACCGCCGACGAGCTCTTCTTCA 3002

Qy    869 TCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACG 928
Db   3001 CCGGCTCCGCCGCCGAAGTCACGCCGATCAAGCAAGTCGACAACGTCGAAATCGGCTCCG 2942

Qy    929 GCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACG 985
Db   2941 GAACCCGCGGCCCGGTCACCGAAGAGCTCCAGACCGCGTTCTTCGACCTCGTGGAGG 2885
```

```
RESULT 9
AP003000/c
LOCUS       AP003000               348077 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 7/21.
ACCESSION   AP003000 BA000012
VERSION     AP003000.2  GI:14023048
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 348077)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994975.
FEATURES             Location/Qualifiers
     source          1. .348077
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
     gene            complement(169. .1065)
                     /gene="mll2551"
     CDS             complement(169. .1065)
                     /gene="mll2551"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transcriptional regulator"
                     /protein_id="BAB49656.1"
                     /db_xref="GI:14023049"
                     /translation="MALDWDKLRVFHAAAEAGSFTHAAETLHLSQSAISRQVSALEHD
                     VGVPLFNRHARGLVLTEQGEMLFRTAHDVLMKLETIKSRLTETKDRPSGVLRVTTTVG
                     LGAGWLTERVQEFIELYPEISLQLILANEELDLTMRQADCAIRLRQPQQPDLIQRRLF
                     TVHFHLYAAPSYVAKHGKPASVSELRNHRIVTFGLPVPSHLSELNWLETVGDFEGGQR
                     VPSLQINDILSIKRAVQGGAGIAMLPDYVINKDSGLVQLLPETEVPSFDTYFAYPDAM
                     KNQAKLHVFRDFIIAKARSWSF"
     gene            complement(1156. .2130)
                     /gene="mll2552"
     CDS             complement(1156. .2130)
                     /gene="mll2552"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin reductase"
                     /protein_id="BAB49657.1"
                     /db_xref="GI:14023050"
                     /translation="MTTKHAPVLIIGSGPAGYTAAVYAARAMLKPMLVAGLQQGGQLM
                     ITTDVENYPGFADPIQGPWLMEQMMKQAEHVGTDIINDIITEVDLNVRPFRAKGDSGT-
                     TYTADALIIATGAQAKWLGIPTEQDFMGFGVSACATCDGFFYRGKDVAVVGGGNSAVE
                     EALYLSNLAKSVTVIHRRSDFRAERILRERLLQKDNVRVIWDTVVDEITGRPGKAPLP
                     PSVEGLKLKHAVTGAETHLKVDGVFVAIGHAPAVELFVGKLKQKPNGYLWTAPNSTRT
                     DVPGVFAAGDVTDDVYRQAVTAAGLGCMAALEAEKYLAGIEVHREAAE"
     gene            2378. .2857
                     /gene="mlr2554"
     CDS             2378. .2857
                     /gene="mlr2554"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transcriptional regulator"
                     /protein_id="BAB49658.1"
                     /db_xref="GI:14023051"
                     /translation="MPLKADLDAIDWKILRELQADGRMTNVELSNRVGISAPPCLRRV
                     KRLEEAGIIRGYRALLNAPALGLDVVAFCLVGLHHQADAELRTFAERTRGWPIVRDAW
                     MVSGESDFLLHCLASDLGAFQTFVIEELTSTPNVDTVRTALTIRRVKDEGLVSFERS"
     gene.           3093. .3887
                     /gene="mlr2557"
     CDS             3093. .3887
                     /gene="mlr2557"
                     /codon_start=1
                     /transl_table=11
                     /product="glucosyl transferase"
                     /protein_id="BAB49659.1"
                     /db_xref="GI:14023052"
                     /translation="MTRFASPVSALIICVNEADMIGPCLESIDFCAEIIIVDSGSTDG
```

```
                TIECVKSYIAKGYPIRLLHNDWPGFPRQRQFSLDHATQPWCLSIDPDERVDDQLRQSI
                VAITQAADDKISGWYIRRRDWLKGYGYAHRWVLHNRLMRLFRREGATIDLTARVHEAF
                QVPGETGTVEDGVLLHRREISVEEDLARANAYSSLKAVTLVEKGKKPGLVRLVLSPLG
                NFLKFYLAKRYFLCGRHGFVYSMSVMIYSPATEAKLYEASERDDPV"
gene            complement(3903. .4556)
                /gene="mll2559"
CDS             complement(3903. .4556)
                /gene="mll2559"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB49660.1"
                /db_xref="GI:14023053"
                /translation="MADKTGTPHPLTEPGVWIERIGGRVFPPHLPALFLDRDGTINVD
                TDYPSDPAEIVLRPQMLPAIATANRAGIPVVVVTNQSGIARGYFGWSAFAAVNGRVLE
                LLREEGVFVDMVLACAYHEAGVGPLAIPDHPMRKPNPGMLVEAGKRLALDLQRSLIVG
                DKLADMQAGKRAGLAQGWLVDGEAAVQPGFAIRPLRDSSELGDLLAAIETLGRDNRS"
gene            complement(4549. .5121)
                /gene="mll2561"
CDS             complement(4549. .5121)
                /gene="mll2561"
                /codon_start=1
                /transl_table=11
                /product="probable phosphoheptose isomerase"
                /protein_id="BAB49661.1"
                /db_xref="GI:14023054"
                /translation="MSELNDYLVRSAAAITAMVERDLTGEMERAASAVVTALSSGKAL
                LIAGNGGSASDAIHIAGELVGRFLKERKAYNVIALPANAAVLTAWGNDYGFDTVFSRQ
                VEAHGTAGGVLLAISTSGNSPSILAAAEQARMMDMTVIGLTGDTGGKLKPLCDILLNV
                PSTSTPIIQQGHLCLYHHLCEVVEARLSNG"
gene            complement(5129. .6619)
                /gene="mll2562"
CDS             complement(5129. .6619)
                /gene="mll2562"
                /codon_start=1
                /transl_table=11
                /product="ADP-heptose synthase"
                /protein_id="BAB49662.1"
                /db_xref="GI:14023055"
                /translation="MIKHNPPSPEPLHRAIARFGGVTVLVVGDLILDRFVNGVIERIS
                PEAPIPVLHGRGETSAMGGAGNVVANIVSLGARAIPVSVIGTDTAGDSLVRMLAELGA
                ETAGLSQQRGRMTSSKSRFSALNQQVLRFDEEEIKPLDETERAGLIRHFRAALAGAEI
                VILSDYGKGILLDGVAAELIAICREAGKPVLVDPKGRDYARYAGATAITPNRKELGEA
                VGHAVFADDEIVAAARELISAHGFDFVVATRSEKGMSVVGPDEARHIATQAREVFDVS
                GAGDTVIATFALALASGADPVAAASIANAAGGVVVGKRGTARLTVEELTGALFRSHGP
                TAHKDAILDASSAARMVAAWKEEGLSVGFTNGCFDILHAGHVSLLHAARSQCDRLVLG
                LNSDASVRRLKGPGRPVNDQHDRACVLAALASVDAVVVFEEDTPLALIEALLPDILVK
                GADYTIDTVVGADVVQKAGGRVVLVDLVAGKSTTGTIGKLRAGSTS"
gene            complement(6734. .7366)
                /gene="mll2564"
CDS             complement(6734. .7366)
                /gene="mll2564"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB49663.1"
                /db_xref="GI:14023056"
                /translation="MPSTELLIAFFATTAIFAYIPGPAMLYAAAQTLARGRWSGLTAA
                LGIHLGGYVHVFAAAAGLSVLFHAVPPLYMAVKLIGALYLIWLGVSLFRKRVEGDVAL
                PAIERKSARRAFFESITVEVLNPKTAIFFMAFLPQFIDASAALPVWLQFVVLGTIVNL
                MFSSADIVCVFLAGLVISRLRRSSRAQRVMQRAGGAVLVGLGVHVALQKS"
gene            7499. .8449
                /gene="mlr2565"
CDS             7499. .8449
                /gene="mlr2565"
                /codon_start=1
                /transl_table=11
                /product="ADP-L-glycero-D-mannoheptose-6-epimerase"
                /protein_id="BAB49664.1"
                /db_xref="GI:14023057"
                /translation="MIIVTGGAGMIGSNIVAALNAEGHDDILVVDDLTDGHKIANLAD
                LSIADYLDKDDFLTRIEAGALGRIEAVFHQGACSTTTEWNGKFMMEVNFAYSKRLLHA
                CQAPRVPFLYASSASVYGGGSEFREEPALERPLNVYAYSKKLFDDYVRRTVFDTDHSQ
                VAGLRYFNVYGPREAHKGAMASVAFHLFNQVERGENPRLFGAYDGFGPGEQSRDFIHV
                GDVADVNLWLWKRGSSGIFNCGTGRAQPFRAIAETVIDTLGKGEIEFIPFPDHLKGSY
                QSFTQADMSRLRAAGYNGQFRTVETGVRDYVEWLKAQRSS"
gene            8422. .9429
                /gene="mlr2566"
CDS             8422. .9429
                /gene="mlr2566"
                /codon_start=1
                /transl_table=11
                /product="ADP-heptose-LPS heptosyltransferase"
                /protein_id="BAB49665.1"
                /db_xref="GI:14023058"
                /translation="MAESPTILVIGPRWVGDMVMAQCLFSALKELYPNAAIDVLAPAW
                AAPLVKRMPEIRQQIDLPLKPGALEFTIRRRFGRLLRGRYDMAYVLPGSWKSALIPFF
                ARIPRRFGHLREMRYGLLTDIVPLPDAVKRRTAQAYFSLAKGGSFRAPHLTVDAGNQA
                ALLDRFGLAPRKFAALMPGAEFGPAKRWPSESYAGLARKFMDKGLKVALFGSKNDRDV
                TAEITALAPGVVDLAGQTRLEDAIDLIAAARLAVSNDSGLMHVAAGVGTPIVAVYGST
                SPENTPPLAERRELVWLHLSCSPCHQKVCPLGHLNCLKTLEVGQVATAAERLLELPAT
                A"
gene            9426. .10400
                /gene="mlr2567"
CDS             9426. .10400
                /gene="mlr2567"
                /note="unknown protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB49666.1"
                /db_xref="GI:14023059"
                /translation="MKVLIVKTSSMGDVIHTFPAVEDASLHRPDISFDWCVEEAFAGI
                VALHPAIAGIHTVAIRRWRKALFNGGTWREAAALRRTLRECRFDLVIDAQGLLKSALV
                ARQARAPIAGFDRSSAREPSATLFYDVTYGVPRNLHAIERTRRLFGLALGYQPDLSTL
                ASGIVPPPGALAGITGKTAFLLHGTSREDKKWPVEDWIGTARLLVERGMTPVTTWSNE
                REKSVAEAIARAVPSTVVVPKSPLADIAAILGRSTLVIGADTGLTHLASAFGLPTVAV
                FLATEPGLTGPRGPYASTLLAVPGGRLTPAEVVAEADSLLNLKLQAPA"
gene            complement(10535. .11008)
                /gene="mll2568"
CDS             complement(10535. .11008)
                /gene="mll2568"
                /codon_start=1
                /transl_table=11
                /product="transcriptional elongation factor"
```

```
Query Match             9.3%;  Score 95;  DB 1;  Length 348077;
Best Local Similarity   48.0%;  Pred. No. 0.00011;
Matches  368;  Conservative    0;  Mismatches  390;  Indels    9;  Gaps    3;

Qy     133 GGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATC 192
           | |||| ||  |||||| | ||||||| | |  ||   |  || || ||||||    |
Db   79113 GACGGCTTCATCTGGATGAATGGCGAATTCGTTGCGTGGGGCGACGCCAAGATCCATGTT 79054

Qy     193 TTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGC 252
           |      ||| |      |  ||      |  || |   | || ||  |  |    ||||
Db   79053 CTGACGCACGGTCTGCACTACGCCAGCGCGGTCTTCGAGGGCGAGCGCGCCTATGGCGGC 78994

Qy     253 AACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGC 312
           | |||||| |||| ||| ||    || ||  ||| ||   |||| |  ||| ||
Db   78993 GAGATCTTCAAGCTCAACGAGCATACCGAGCGCCTGCATGAATCGGCGCGCCTGCTCGGC 78934

Qy     313 CTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATG 372
           | |       ||       | ||| ||   |||   | | |  | |    ||  || |
Db   78933 TTCAAGATTCCCTATTCGGTTGCCGAGCTCAACGACGCCTCGACCACGCTGTTGAAGAAG 78874

Qy     373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
             | | |  || | |  || | | ||   || || || |   | |  ||| | |
Db   78873 CAAGGGTTCCAGGACGCCTATGTCAGGCCGA---TCGCCTGGCGCGGCAGCGAGCAGATG 78817

Qy     433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
           ||||    ||  | |  ||| | | |    | ||  |||| | |  |   | || | |
Db   78816 GGCGTTTCGGCGCAGAACAACCGTATC---AACGTCGCCATCGCCATCTGGCAGTGGCCG 78760

Qy     493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
           |   |||  || ||  || || |    |||| | |      ||  ||       |
Db   78759 AGCTATTTCGATCCGGCGCAGAAGCTGAAGGGCATCCGCCTCGATGTCGCGGAATGGCGC 78700
```

```
Qy     553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
           || | || |   ||||   |           | |       |  ||  |          |
Db   78699 CGGCCCGACCCGCGCACGGCGCCGTCGAAATCGAAGGCCGCCGGCCTCTACATGATCTGC 78640

Qy     613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
           |||    | |    ||  ||| | |       ||       |  ||| ||  |||||||| |
Db   78639 ACCATGTCAAAGCACGCCGCCGAAGCAAAAGGCTACGCCGACGCCATGATGCTCGACTGG 78580

Qy     673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAG--- 729
           |  |   || |||||||   | |||  |||| | | | ||||||||||||||||
Db   78579 CGCGGCCAGGTCGCCGAAGCAACCGGCGCCAACATCTTCTTCGTCAAGGACGGCAAGATC 78520

Qy     730 CTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTG 789
           |   |  ||  | ||     |    |    ||||||||||    |||||  |||  ||
Db   78519 CATACGCCCAAGCCCGATTGCTTCCTCGACGGCATCACCCGCCGCACGGTCATCGGGCTC 78460

Qy     790 GCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGAC 849
           |||    |    ||| | ||||  | |   |||| | ||          ||||| | | |
Db   78459 GCCAGGGATCGTGGCCTGGAAGTGATCGAACGCGCCATCCTGCCGGAAGAACTCGAAGGC 78400

Qy     850 GCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGAT 896
           |||  | |   |   |  |   |||||||  |  || || ||| |
Db   78399 TTCGAACAATGCTTCCTGACCGGAACCGCGGCGGAAGTAACGCCGGT 78353


RESULT 10
AL646060/c
LOCUS       AL646060              212050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 chromosome, complete sequence;
            segment 4/19.
ACCESSION   AL646060 AL646052
VERSION     AL646060.1  GI:17427566
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 212050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .212050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
     gene            59. .757
                     /gene="RSc0556"
                     /note="RS04909"
     CDS             59. .757
                     /gene="RSc0556"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD14086.1"
                     /db_xref="GI:17427567"
                     /translation="MALTSSAFWFALGSIILTNIVLSGDNAVVIALAARNLPRRQQRQ
                     AIFWGSAAAIALRIALTVLAVKLLALPYLKTIGAVLLVYIGIKLLSEAEDEAAGEHRP
                     HDGLWPAIRTILIADLVMSLDNVVAVAAAAEKGPPGTDVALLVLGLGLSIPLIIFGST
                     LLVALMVRLPLIVTLAAALLGYLAGDMLVTDPVDAGWFMQAIPYADVAIGCLGALIVV
                     IVGWWMGRRSLRQA"
     gene            951. .1472
                     /gene="pilA1"
                     /note="RSc0557; RS04908"
     CDS             951. .1472
                     /gene="pilA1"
                     /function="structural elements; cell exterior; surface
                     structures"
                     /note="Product confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="PROBABLE TYPE 4 FIMBRIAL PILIN SIGNAL PEPTIDE
                     PROTEIN"
                     /protein_id="CAD14087.1"
                     /db_xref="GI:17427568"
                     /translation="MASKHSRNRRRGQRGFTLIELMIVVAIIGILAALAIPAYQDYTI
                     RSRVVEGLALAGQAKALVVENAANAQSSLSLGSASLPASKNVSALNIDSSTGEIAVVY
                     APAVTEAGANTLVLTPYSGSSSVFGNLQAGSTPPAMVLWVCAAAGKSMPVNTVTQRTT
                     ATLPPKYAPSECR"
     gene            1739. .2245
                     /gene="pilA"
                     /note="RSc0558; RS04907"
     CDS             1739. .2245
                     /gene="pilA"
                     /function="structural elements; cell exterior; surface
                     structures"
                     /note="predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="TYPE 4 FIMBRIAL PILIN SIGNAL PEPTIDE PROTEIN"
                     /protein_id="CAD14088.1"
                     /db_xref="GI:17427569"
                     /translation="MKSMRHLNKRVQKGFTLIELMIVVAIVGILAAIAIPAYQDYTVR
                     ARVTEGLSLAAQAKALVSENAANAQSDLSVGSSVFTPTKNVANLTIAGTGTVPGQITV
                     TYTTAAGGGTLALVPTAAGTALPVSSAPSGPIMWTCYAQNKAQAASSVAPSGTMSLAA
                     KYAPAECR"
     gene            2351. .4060
                     /gene="RSc0559"
                     /note="RS04906"
     CDS             2351. .4060
                     /gene="RSc0559"
                     /function="miscellaneous; hypothetical/global homology"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD14089.1"
```

```
                /db_xref="GI:17427570"
                /translation="MLWPVWLATVACWSVPFLVAVHSYPIPTFYSEFVAAIGWVALAA
                GVLGSTWHSKAGLPKVVLAPLALIGVLIVQLVVATPLNPFFTFAAIVFLLGTVAVCGL
                GARCRDVPGVLESIAVAVIIGGLLTVAIECLHLFRVSGLPINWISIMPTGAGRRMWGN
                LNQPNHVATYLAFGLAACLFLGSTRRRYWAPLAAIALALLLGMALTVSRMSWLHLVLV
                GGVAGLAWSAEERGARRWIRAGVPVLGLAVVYQLCNWLVAYANVLWHLDLPISLDERL
                QQGVGLRVFLWKHAWHMFLAHPWLGGGWGDYAWNQYVQTDVLGHVEMSMNAHNLVLDL
                LAKVGVFGLLAVMLPFLGLVHAAWKRRMTPALAFLYAVILVTVAHSMLEYPLHYLYFL
                LPFAFALGYVDDRKSRVLSPDTAWVLTGIVAICGAVLTGRMWIDYQSVERLYYSLDGP
                VELQRYQRSGQLLLVPYGNLSIANNAGMTAETAPIMAAVEHQAVQFYPGSGTVQRWAI
                ALAFQGKTDEALTQVRRLHNQYWIDYAGDSKLLTLVCTKKLEGLATFCARLKAENLVV
                GVD"
gene            complement(4289. .4765)
                /gene="moaC"
                /note="RSc0560; RS04905"
CDS             complement(4289. .4765)
                /gene="moaC"
                /function="small molecule metabolism; biosynthesis of
                cofactors, carriers; molybdopterin"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE MOLYBDENUM COFACTOR BIOSYNTHESIS
                PROTEIN C"
                /protein_id="CAD14090.1"
                /db_xref="GI:17427571"
                /translation="MPHLTHFDTAGQAHMVDVGDKAVTHRVAVATGLIHMLPTTFALV
                RDGTAKKGDVLGIARIAAIQGAKRTSDLIPLCHPLALTKVAVDFELDEADRTVRCTVR
                AETHGQTGVEMEALTAVQVGLLTIYDMCKAVDRGMVIGDVRLMEKRGGKSGEWVAQ"
gene            4859. .6544
                /gene="RSc0561"
                /note="RS04904"
CDS             4859. .6544
                /gene="RSc0561"
                /function="miscellaneous; hypothetical/global homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="HYPOTHETICAL SIGNAL PEPTIDE PROTEIN"
                /protein_id="CAD14091.1"
                /db_xref="GI:17427572"
                /translation="MTLSFFRPRARKVTAAVLLMTWLAPVSPLALAQPASVPAAPAGA
                AAAEQVQSNLNRSVQIGRQSPYLQKDATIIDAPTYELPDMGDPSTAALSPEMERRLGD
                RVMRQIRRDPDYVPDPLLSDYLNDIGYRLIEAARRQHVAGSTSASSFELFAVRDPGIN
                AFALPGGYIGVNTGTLVATESESELASVLGHEIGHVLQRHIARGIDKSGESMWIALAS
                ILLAGLAATKSGDAAQALAVGGQAAAVSNQLAFSRGAEREADRVGFTLLTGAGYNPDG
                MPDFFRRLQRVSSIADTGAVPGYARTHPLTGERIADMEDRARGLPHPRQPQRPAFGFA
                KARARVLQETSTSAYLDVRNAMRSQLASAPDAPAEKRAALWYGIAVAEQLAGQLDPAE
                QALLEARRLYGNIPGLTSGSPSLDVTAIELARARNRVPEALTMARAALAAYPLSRAVG
                ITYAQTLLAAGRAGDAIPYLQDKAREDTAQPIWWDMLARAYADQGKRVEQHRALAEKY
                ARDGAWGAAVEQLKLARDAGDADFYTLSEVDARLHQMERQYREEKQEEKALPK"
gene            complement(6559. .7146)
                /gene="RSc0562"
                /note="RS04903"
CDS             complement(6559. .7146)
                /gene="RSc0562"
                /function="miscellaneous; unknown"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="HYPOTHETICAL PROTEIN"
                /protein_id="CAD14092.1"
                /db_xref="GI:17427573"
                /translation="MDEIVRQAMARWPDVPDCFGWLALDRRGQWRMRNEYAQQHRLSG
                DPVRHPALIDFIVRNYTHDAAGRWFFQNGPQRVFVELDCTPWIVRLSPEGAPTALATT
                TGAAFVPAGCFVDEHGNVLLAGHVAGVAGRETLALLHDHDLEPFSSLAHWHGQACGAA
                LGMLPWGDRTFDIQPIRSDEAERRFGFVRHPAALA"
gene            complement(7139. .8254)
```

```
Query Match             8.5%;  Score 87.2;  DB 1;  Length 212050;
Best Local Similarity   53.7%;  Pred. No. 0.0015;
Matches  204;  Conservative    0;  Mismatches  173;  Indels    3;  Gaps    1;

Qy     619 GCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAAC 678
           ||||     |||||||  | ||  ||           ||  | |||||||||| | ||   |
Db   10489 GCCAACCAGGAAGCCACCGGCCTGGGCTACGATGAAGCCCTGCTGCTCGACACCGAGGGC 10430

Qy     679 TGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCC 738
           |  |||  ||||||  |||||    |||||  | |||||||    |||||  | |   | |
Db   10429 TATGTGAGCGAAGGCTCGGGCGAGAACGTCTTCATCGTCCGCAACGGCGTGATCTACACG 10370

Qy     739 CCGTCCCGGAACGCGTTGCCG---GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGAC 795
           ||   || |  | |||  | |   ||||||||||  |  |  |  ||    | || | |
Db   10369 CCCGACCTGGCCTCGTGCCTGGACGGCATCACCCGCGACGCAATCATCACCATCGCGGGC 10310

Qy     796 CAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGAC 855
           |  | ||||||| |    |  |     | ||||| ||   ||  | |||  ||| |||
Db   10309 GACCTCGGCATCGAGGTGCGCGAGAAGCGCATCACCCGCGACGAGATGTACTGCGCGGAC 10250

Qy     856 GAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAG 915
           ||     |      |  |||||||  |  ||||| ||||||  |   |||||||  |
Db   10249 GAAGCCTTCTTCACCGGCACCGCGGCGGAAGTCACCCCGATCCGCGAACTGGATGACCGC 10190

Qy     916 GCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCG 975
           |   | |||   |||  || ||| ||  | ||   |    | | |     |||||  |
Db   10189 GTGATCGGCGCGGGCTCGCGCGGCCCGATCACCACGCAGCTGCAGTCGACGTTCTTCGAC 10130

Qy     976 CTGATGGACGAGCCGGGCCC 995
           |  |  |||    | || |
Db   10129 ATCGTCAACGGCAAGAGCAC 10110


RESULT 11
AE004913/c
LOCUS       AE004913               17576 bp    DNA     linear   BCT 30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 474 of 529 of the complete
            genome.
ACCESSION   AE004913 AE004091
VERSION     AE004913.1  GI:9951286
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 17576)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
   PUBMED   10984043
REFERENCE   2  (bases 1 to 17576)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
```

```
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .17576
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            complement(111. .1067)
                     /gene="PA5001"
     CDS             complement(111. .1067)
                     /gene="PA5001"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG08386.1"
                     /db_xref="GI:9951287"
                     /translation="MKVLFLVQKEQRAILDRLYDGIAAHCECDTRWLSSEEQADLRGY
                     FRKHVDVSRYDRILFFLRFKKEMRQVRFIRSVPNLVILEHDAYQNYIPCKYTGKFSAH
                     YRRLPWARVISSGHTVSERLRQEGFDAVFVPKGYDQTLLHDLGLARDIELGFVGSTGS
                     VAYSGRKALLDELGGVENLLVTKTKSGEEYLRTLNRIRFFVSADVGMGEYMIKNFEAM
                     ACGCVLLAYDQGERENEALGFRDMENIVLYRSVADIQEKLAILRLDTALAERIARSGQ
                     ALAIERFSFAAVGQSIVEAMRPALREPERPGWLEKTWRRIFT"
     gene            complement(1140. .2558)
                     /gene="PA5002"
     CDS             complement(1140. .2558)
                     /gene="PA5002"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG08387.1"
                     /db_xref="GI:9951288"
                     /translation="MSARKQQLLKRHRRNKRIGLLVALLALLAVGLLVSPWLLPILLV
                     ALWVAHEAWFADHLFYSPGEDYRYRFAEGVESLPVRLADGRLRVDGELREGDTLVLGI
                     GVRAGWLGRFLEPSVLLEGGAEADAQAFERGVNGLRYLNLTGLAGPLGEGRIRLRGRH
                     CRLVGEPTLWRARHPDYRDRRVMVIAPHADDAELAAFGLYSQAREAWIVTLTAGEIET
                     EHYRRMGLDGIAAARLKGRLRAWDSQAVPTWGGVPAERCVQLGYFCLQLPAMQANPGE
                     VVPSREADLADIRPFRQFNRLRLASDADGLSTWNNLLADLRELILLARPEVIVLPHPH
                     FDPHPDHVRAQEAVREALQGLDWQPQALLHYANHLHDNDRWPMGDAHMGVSLPPLTEE
                     CSPLLPWTLALERTRQVDKAMALGMMHDLQPRPPFKRRLRRRLQGWLAGRRWPAYGED
                     EFMRKAVRRHELFWVESLDGEA"
     gene            complement(2562. .3458)
                     /gene="PA5003"
     CDS             complement(2562. .3458)
                     /gene="PA5003"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG08388.1"
                     /db_xref="GI:9951289"
                     /translation="MLNRLRAFRERGWREVDAPTYAQAWQRFGGSVATHPLVVEQLAE
                     MAGIPVRYLGWESAGELVAAIPTWGRHLALSKDVLKRRGKKGLYDLGNAELILPVAGD
                     ARIALRHAGRYLSALHERNIVTLRPQTEQLAFARAPEDFSKKFRYNQRREVRLVEEAG
                     GQVRQADSFAPAEFAAIYLELFQRRWGFPATGAERMAEVFERLRELLRGSVLLLNDQP
                     IAVQVLYRAEAPGWLSVEYINGGVDPQTREFSPGSVLSYLNTQAAWEDARALGKPLRY
                     SFGRADREYKDRWCHPSPVFQV"
     gene            complement(3452. .4588)
                     /gene="PA5004"
     CDS             complement(3452. .4588)
                     /gene="PA5004"
                     /codon_start=1
                     /transl_table=11
                     /product="probable glycosyl transferase"
                     /protein_id="AAG08389.1"
                     /db_xref="GI:9951290"
                     /translation="MTRSAEPRVLQFCHGYDGPFLDCARQYASLFAGTPYKVTTVFLT
                     GRSDPEVAARCASDEVLFLEYGSRDIRGLKLKAIRDLREIARSRDFRFCIAHRFKPIW
                     IASLATRLPIVGVHHAFGDYQRLSRKVFAGLMSKRLSLLAVSDAVRDDIRRCLPKWPA
                     ERIETLYNRIDVAALQAEQVERLSARRQLRLPDEAWVVGNVGRLHPDKDQATLLRGFA
                     AALPRLPQNSLLAILGSGRLEEQLKDLACELGIGERVLFLGQVEEARRYFKAFDAFAL
                     SSDHEPFGMVLLEAMVAGVPLIATACGGAREVVEGVGILFPLGDETALAEGLTHLAAL
                     DRRQREACARLMLERLETRFSDAAVRREFWRLPTIAGFASESAC"
     gene            complement(4572. .6329)
                     /gene="PA5005"
     CDS             complement(4572. .6329)
                     /gene="PA5005"
                     /codon_start=1
                     /transl_table=11
                     /product="probable carbamoyl transferase"
                     /protein_id="AAG08390.1"
                     /db_xref="GI:9951291"
                     /translation="MALTVLGLSGAVSHDPSAALYIDGKLVAAVEEERFVRDKHAKNR
                     MPYESAKFCLEQAGIEPADVDVVAIPFAPISIMEKARWHYAKRYAYAPDRALDAILLG
                     NRRYKRYYKRIEWCLQQLGFDLKKIKIQPVEHHLAHASSAYHCSGFKEKTAILGIDGK
                     GEYATTFFGWGENGRIHKIKEFYDPDSLGGLYGAITEYLGFDMLDGEFKVMGMAPYGD
                     ASKYDFSRLAKFENGELVINTDYANVIGFRRYKEKGKGYYFSPKLIEWLGPKREGDIA
                     DDPYIHYAASIQALFEKLALEMMDYYLGDIIRETGKVAFAGGCALNVKLNQKIISRPE
                     VKELFVQPAASDAGTAVGAAAYVSHQRGVPVEKMEHVYLGPEYSNEDVIAACARHEAR
                     PQWQKIHNRSRRIGRILADGNPVAWFQGRMEFGPRALGGRSILGCPSVPGVADRINAQ
                     IKFRERWRPFCPSMLDTVGPQMFKIDHPAPFMTFTFEVNEEWKERVPEVVHEDGTSRA
                     QVLKREYNPRYYDLMKELENLTGNGVALNTSLNRRGEPMICSPTDALNMFFGSDLQYL
                     IMEDILVVKEGVDWYDSVG"
     gene            complement(6433. .7911)
                     /gene="PA5006"
     CDS             complement(6433. .7911)
                     /gene="PA5006"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG08391.1"
                     /db_xref="GI:9951292"
                     /translation="MRLAELQAAGRQPQLPLRVTLADGAGAAELRLDSLLRVLPGQRY
                     VGAGEWRGRKVLAKLLVGGKAARHFQRELEGARLLAAQGLTTPALLADGLQQGEGGWL
                     LFDYLDAAQSLGDAWRAVEAQAPLGDAQQAVLGEALAAIGELHAKGLWQDDLHLDNLL
                     RHAGRLYLIDGGGVRAETPGQPLSRQKVLENLGVFFAQLPALLDPFLEELLVHYLLAN
                     GEHALPLEALQKEIAKVRRWRVRDLLRKVGRDCSLFSVVRGPFRLFAARRSEAETMAP
                     LLADADRLIAAGHVYKDGGAATVARVEANGRTLIIKRYNIKGWVHWLKRFWRPSRAWA
                     SWREGNRLEFLGIATARPLAMLERRWCWLRGRAYLITDYLGGQDIIARFKPCLDMPDG
                     AVSPPEAELAALDRLFSTLIRERISHGDLKGHNLFWQAHEQGGEWALIDLDAMRKHHS
                     QASFARAYARDRARFLRNWPQDSGLYRLLEQRIPRLGDAAGA"
     gene            complement(7908. .8666)
                     /gene="PA5007"
     CDS             complement(7908. .8666)
                     /gene="PA5007"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG08392.1"
                     /db_xref="GI:9951293"
                     /translation="MSVFIAAQDHDLLQRNRLDSFDALWALQLDAVDEPNTERGGWSS
                     VYRLDLGDSAYYLKRQSNHLTRSLRHPLGEPTFAREFRNIERYRALGIPALQAAFFAQ
                     QAVGGERQAILLTRALDGWQDLDTWLASWHDLAEERRRAILRAVGRLARQLHAAGQMH
                     GCFYPKHVFLREGADGVFEACLIDLEKTRPIWFGRRDQVKDLEPMLRRAPDWSEADVH-
                     EFISAYVDEPLDSAEVEQWYRLLDQRRRRKESRR"
     gene            complement(8663. .9397)
                     /gene="PA5008"
     CDS             complement(8663. .9397)
                     /gene="PA5008"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG08393.1"
                     /db_xref="GI:9951294"
                     /translation="MAGWTLEPDYRHLAADFGSLDSVFALEGERLTRDPLSEVIRIER
                     DGVRYYVKRYRSAGKGLRRYLARPRIKAEWQNLKRFEKWGIPTAEVVAWGLERKAGAF
                     QRGAMITRELPGTEDLSVLACNRDPRLKDPRWVDGVSRQIAKATRLMHDQHFTHNDLK
                     WRNLLVDDQPLVYFIDCPNGAFWWSFMLRYRITKDLACLDKVAKYHLSRTQRLRFYLQ
                     YRRRRHLNESDKKRIRQVVAFFEGRE"
     gene            complement(9397. .10203)
                     /gene="waaP"
                     /note="PA5009"
     CDS             complement(9397. .10203)
                     /gene="waaP"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="lipopolysaccharide core biosynthesis protein
                     WaaP"
                     /protein_id="AAG08394.1"
                     /db_xref="GI:9951295"
                     /translation="MRLVLEEPFKRLWNGRDPFEAVEALQGKVYRELEGRRTLRTEVD
                     GRGYFVKIHRGIGWGEIAKNLLTAKLPVLGARQEWQAIRRLHEAGVATMTAVAYGERG
                     SDPARQHSFIVTEELAPTVDLEVFSQDWRERPPPPRLKRALVEAVARMVGDMHRAGVN
                     HRDCYICHFLLHTDKPVSADDFRLSVIDLHRAQTRDATPKRWRNKDLAALYFSALDIG
                     LTRRDKLRFLRTYFRRPLREILRDEAGLLAWMERKAEKLYERKQRYGDLL"
     gene            complement(10200. .11321)
                     /gene="waaG"
                     /note="PA5010"
     CDS             complement(10200. .11321)
                     /gene="waaG"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose:(heptosyl) LPS alpha
                     1,3-glucosyltransferase WaaG"
                     /protein_id="AAG08395.1"
                     /db_xref="GI:9951296"
                     /translation="MTLAFILYKYFPFGGLQRDFMRIALECQRRGHDIRVYTLIWEGD
                     VPDGFEVLVAPVRSIFNHRRNEKFTAWVRADLDRRPVQRVIGFNKMPGLDVYYAADAC
                     FEEKAQTLRNPLYRQWGRYRHFAGYERAVFDPASKTEILMISEVQQPLFVKHYGTQAE
                     RFHLLPPGISQDRRAPANAADVRAEFRREFGLEEDDLLLVQIGSGFKTKGLDRSLKAL
                     SALPKALRRRTRLIAIGQDDPKPFLLQIAALGLNDQVQILKGRSDIPRFLLGADLLIH

  Query Match             8.4%;  Score 85.8;  DB 1;  Length 17576;
  Best Local Similarity   48.2%;  Pred. No. 0.0041;
  Matches  336;  Conservative    0;  Mismatches  352;  Indels    9;  Gaps    3;

Qy    256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
          |||||||| ||    |  ||    |||||| || |||||   |||  ||   ||  | ||
Db  14249 ATCTTCCGCCTGCAGGCGCATACCGACCGGCTGTTCGACTCCGCGCACATCATGAACATG 14190

Qy    316 GACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCG 375
          |   |  |||||||   |||||  |   ||||   |||         ||||     |
Db  14189 CAGATCCCGTACAGCCGCGACGAGATCAACGAGGCGACCCGCGCCGCCGTGCGCGAGAAC 14130

Qy    376 CAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGC 435
          | |||       | |   |    | ||  ||   |   |||  ||   |  ||| ||||
Db  14129 AACCTGGAAAGCGCCTATATCCGCCCGATGGTGTTCTACGGAAGCG---AAGGCATGGGC 14073

Qy    436 GAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGG 495
          |   | |      |  |     |||| | |   |  |  |  | |  |  |   |
Db  14072 CTGCGCGCCAGCGGCCTGAAGGTCCATGTGATCATCGCCGCCTGGAGCTGGGGCGCCTAC 14013

Qy    496 GCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGC 555
          | |   |||  ||||      |||| ||  | |  |         |     ||||
Db  14012 ATGGGCGAGGAAGCCCTGCAG---CAAGGCATCAAGGTGCGCACCAGTTCCTTCACCCGC 13956

Qy    556 CGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACC 615
          | |  || |  || | |  |   ||   |||     |||  |   |       |    |
Db  13955 CACCACGTCAACATCTCGATGACCCGCGCCAAGTCCAACGGCGCCTACATCAACTCGATG 13896

Qy    616 GCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGAC 675
            |||      ||||| |    | | || ||    |   || ||  |||||||  ||||
Db  13895 CTGGCCCTCCAGGAAGCGATCTCCGGCGGCGCCGACGAGGCCATGATGCTCGATCCGGAA 13836

Qy    676 AACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCC 735
          || ||||||||||||  | |||   ||| | | |||| |||||| |||  | |   |
Db  13835 GGCTACGTGGCCGAAGGCTCCGGCGAGAACATCTTCATCATCAAGGATGGCGTGATCTAC 13776

Qy    736 TCCCCGTCCCGGAACGCGTTGCCG---GGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
          |||||      | ||| |  | |   |||||||| ||||| ||  |  |    ||||||
Db  13775 ACCCCGGAAGTCACCGCCTGCCTGAACGGCATCACTCGTAACACTATCCTGACCCTGGCC 13716

Qy    793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
          | | |    ||  | ||     |  |     | ||||| ||   ||  | |||  ||||
Db  13715 GCCGAACACGGTTTTAAACTGGTCGAGAAGCGCATCACCCGCGACGAGGTGTACATCGCC 13656

Qy    853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
          ||||||    |         ||| || |  |  ||||| ||||||  |    |||| ||

Db  13655 GACGAGGCCTTCTTCACTGGCACTGCCGCGGAAGTCACGCCGATCCGCGAAGTGGACGGT 13596

Qy    913 GAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGG 949
          |  |||  ||||   |  || ||  | || |
Db  13595 CGCAAGATCGGCGCCGGCCGCCGTGGCCCGGTCACCG 13559


RESULT 12
AE011153
LOCUS       AE011153               10993 bp    DNA     linear   BCT 03-APR-2002
DEFINITION  Methanosarcina acetivorans str. C2A, section 498 of 534 of the
            complete genome.
ACCESSION   AE011153 AE010299
VERSION     AE011153.1  GI:19918472
KEYWORDS    .
SOURCE      Methanosarcina acetivorans C2A.
  ORGANISM  Methanosarcina acetivorans C2A
            Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
            Methanosarcinaceae; Methanosarcina.
REFERENCE   1  (bases 1 to 10993)
  AUTHORS   Galagan,J.E., Nusbaum,C., Roy,A., Endrizzi,M.G., Macdonald,P.,
            FitzHugh,W., Calvo,S., Engels,R., Smirnov,S., Atnoor,D., Brown,A.,
            Allen,N., Naylor,J., Stange-Thomann,N., DeArellano,K., Johnson,R.,
            Linton,L., McEwan,P., McKernan,K., Talamas,J., Tirrell,A., Ye,W.,
            Zimmer,A., Barber,R.D., Cann,I., Graham,D.E., Grahame,D.A.,
            Guss,A., Hedderich,R., Ingram-Smith,C., Kuettner,C.H.,
            Krzycki,J.A., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B.,
            Reeve,J.N., Smith,K., Springer,T.A., Umayam,L.A., White,O.,
            White,R.H., de Macario,E.C., Ferry,J.G., Jarrell,K.F., Jing,H.,
            Macario,A.J.L., Paulsen,I., Pritchett,M., Sowers,K.R.,
            Swanson,R.V., Zinder,S.H., Lander,E., Metcalf,W.W. and Birren,B.
  TITLE     The Genome of M. acetivorans Reveals Extensive Metabolic and
            Physiological Diversity
  JOURNAL   Genome Res. 12 (4), 532-542 (2002)
  MEDLINE   21929760
   PUBMED   11932238
REFERENCE   2  (bases 1 to 10993)
  AUTHORS   Birren,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) Center for Genome Research, Whitehead
            Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
FEATURES             Location/Qualifiers
     source          1. .10993
                     /organism="Methanosarcina acetivorans C2A"
                     /strain="C2A"
                     /db_xref="taxon:188937"
     gene            78. .629
                     /gene="MA4346"
     CDS             78. .629
                     /gene="MA4346"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein (multi-domain)"
                     /protein_id="AAM07689.1"
                     /db_xref="GI:19918473"
                     /translation="MQENPEFEEKKEPKESDFKAPGELETGYTEKSWEKPEKTMEEKE
                     AGIEPTPQERIEPKESDFKAPGELETGYTEKSWEKPEKTEEEKEAGKETGEEAGKETG
                     EEERTVYTMKKRLTKSKSDRMLFGVCGGLGKYFGIDPTFVRLLFVLLAFVNGIGIVLY
                     IILAIIMPSEESVQMVPETRKSS"
     gene            complement(716. .1255)
                     /gene="menG"
                     /note="MA4347"
     CDS             complement(716. .1255)
                     /gene="menG"
                     /codon_start=1
                     /transl_table=11
                     /product="menaquinone biosynthesis methyltransferase
                     (2-heptaprenyl-1, 4-naphthoquinone methyltransferase)"
                     /protein_id="AAM07690.1"
                     /db_xref="GI:19918474"
                     /translation="MGRFLDSDFRRKLQSPDKLIDRSGIKEGMHVLEVGCGSGAFTTF
                     VARTVGIKGEVYALDIQPGMLMQLKEKLSRPENRDIRNIKLIKGDAHNLPFDDNSFDL
                     VYAITVIQEIPDKNKVLKEIKRVLKPGGILAVTEFLPDPDYPLKSTTIRLGEEAGLIL
```

```
                DKVEGNLWHYTVRFKTKKK"
gene            1795. .2088
                /gene="MA4348"
CDS             1795. .2088
                /gene="MA4348"
                /codon_start=1
                /transl_table=11
                /product="predicted protein"
                /protein_id="AAM07691.1"
                /db_xref="GI:19918475"
                /translation="MEIFSGRQCYGEGYTHPDPFDLKSLKAEPATLEELSMRHYGDES
                KEALQGEDERKQLHWYGSCSPYVTAPGSFCIARLNTASSADLCRVACFLFVNK"
gene            3071. .3955
                /gene="ilvE"
                /note="MA4349"
CDS             3071. .3955
                /gene="ilvE"
                /codon_start=1
                /transl_table=11
                /product="branched chain amino acid aminotransferase"
                /protein_id="AAM07692.1"
                /db_xref="GI:19918476"
                /translation="MQMSDLLIYLNGKFVPKDEAKVSVYDHGFLYGDGVFEGIRAYNG
                RVFKLKEHVDRLYDSAKAIAMDIPVTREEMTEIILEALRKNNLKDAYIRPIVSRGIGD
                LGLDPRKCGKPSVIVIAQGWGAMYGDLYEVGLTGVSVCVRRNAPDALSPNIKSLNYLN
                NILAKIEANEKGGDEAIFLDHNGFVCEGSGDNIFVVKDDRVLTPFTISNLKGITRATA
                IELLDEMGYKVIESNLGLFDLYTADEIFVTGTAAESAPVTRLDGRVIGTGKPGPLTMK
                MVEAFAKITQSTGTQIYK"
gene            4274. .6214
                /gene="moeA"
                /note="MA4350"
CDS             4274. .6214
                /gene="moeA"
                /codon_start=1
                /transl_table=11
                /product="molybdenum cofactor biosynthesis protein"
                /protein_id="AAM07693.1"
                /db_xref="GI:19918477"
                /translation="MKRKEFRELVPVEEARRIINRLQIRPEKENSALENAYGKILAED
                IVTEINVPPFSRATMDGYAVRAEDTYACSETEPVKLKLLGNIPAGSDVSFKVSAGEIV
                EIATGAPIPEGADAVVMVEYSSEENGTVLIFRPVAGGENVMKAGSDILKGGRVLRRGR
                KLGTREIGVLASIGKKEVPVLRLPVGIISTGNELVSPGGNLAPGKVYDANSYTLYTGV
                QECGAFSLLYGVVKDDETGIRKALETAVSECALILTSGSTSSGTGDVMYRIIDEVGET
                LAHGINIKPGKPVVIGIIKGVPIIGLPGNPSSALMIFNEFVAPFLRKALGAEAGVRRT
                ETGILGTVLRSEGRQQLLPVGLVRGRVYPADRGSGAITSLSEADGFIEIPPETEYMEA
                GTAVEVTLFGEVEKPDLLIAGGFCPGLDVLEDLSGLRFRTLYTGSSGGFSAIAAGTAD
                IAGVNMPSKRPENRRESLPGEIRYNAPTIENMGLSEVVLVKGYRREMGLLVQQDSPVS
                GLEALPGKRLINRNRGSGTRALLELKIEELAEKKGISKKEFTDSIPGYGSGAKSEVSV
                CEAVLSGKVDAGVGLRNCAERNGLKFVKFAEEEYDFLIRKEVLDVPEVRKFLKTLSSA
                EFASKLPAGLQVYERTGEIISY"
gene            7373. .8467
                /gene="MA4351"
CDS             7373. .8467
                /gene="MA4351"
                /codon_start=1
                /transl_table=11
                /product="GTP-binding protein"
                /protein_id="AAM07694.1"
                /db_xref="GI:19918478"
                /translation="MSSIQEQIQEVEDEIRKTQYNKATSHHIGRLKAKIARMRDEIEK
                KASSKGGGEGYSVKKSGDGTVTLVGFPSVGKSTLLNKVTGAKSEVGAYEFTTLTVVPG
                VLEHKGATIQFLDVPGLVKGASSGRGRGKEVISVIRNSDMVIFLLDVFQPKHYEVLMD
                ELYQAGIRVDEVPPDVTIKRKDRGGIEINSTIDLDLDEETIKAVLDEYKIHNAHVLIR
                ENITVDQLIDVVLGNRSYVRSLTVVNKVDLAYPQLIEECRKLYPKSIFISAHEGINIE
                NLKDSIYDCLGFIRVYLKPQGQPADMEEPLIVMSGTNIGQICDRLHRDFRRKFRYAQV
                WGPSAKHPGQRVGLEHIMEDEDILMIIIQK"
gene            complement(8761. .10872)
                /gene="MA4352"
CDS             complement(8761. .10872)
                /gene="MA4352"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein (multi-domain)"
                /protein_id="AAM07695.1"
                /db_xref="GI:19918479"
                /translation="MTQNTTDSNPRRKRAKGTVQTKTLGPVDDLEPHVRSDWWQTLFN
                SLYLKTDADLLDDMEVTRKEADLLVSILGLDPEDMVLDLCCGQGRHVLELARRGFSNV
                EGYDRSQYLIRKARTRAQKENLQVRFREGDARKLPYPSDTFSVVTILGNSFGYFDSTL
                HDRKVLEEVFRVLKSGGRVFIDAVDGDHMKKNFQPRSWEWLDRKYFVCRERALSSDGD
                RLICREVICRNDRGIIADQFYAERLYNRESLFELLTASGFSIPTFHTTFSPVSTGTQD
                AGMMGQRILLSATVEKAWPSLSSSLQNAESKKPLNVVVVLGDPRKEDQVKPACVFDDD
                DLETVNRMKQALSEIPYMKFTFLDRHETLLEDLKKRAGKTDLVLNLCDEGFNNDPIKE
                LHVPALLEQFNIPYTGAGPQCLSFCYDKSLVRGAAREMRIPVAKGVLVTGDSDVSSFS
                LSFPLLVKPNSGDSSYGITQKSIVHSREEIFDIMKETREKIGADKPFLLEEFLPGKDI
                SVGIIGNPPFCTVLPITEEDYSAVPEDLPRICGYEAKWLPDSPYWKIKSVPAGLPEKT
                EKEIVRCCLALFTRLGCRDYARFDWRLDAEGRPKLLEVNPNPGWCWDGHLAKMAAYAN
                ISYSGMLAAILEAAKKRCGIGTSVKIELQRKASERSPRKSPDEYAAEFEEEVEAKGPI
                ESENDRKRNVFSGNSQTMQAL"
BASE COUNT     3241 a    2382 c    2591 g    2779 t
ORIGIN

  Query Match             8.2%;  Score 84;  DB 1;  Length 10993;
  Best Local Similarity   45.9%;  Pred. No. 0.0081;
  Matches  371;  Conservative    0;  Mismatches  425;  Indels   12;  Gaps    2;

Qy     152 AGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTC 211
           | ||  |||   | ||    || ||||| ||  | ||  |||  |||  ||| |||  |
Db    3102 ACGGGAAATTTGTCCCCAAAGACGAAGCCAAAGTTTCGGTCTATGACCACGGCTTCCTTT 3161

Qy     212 ATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCG 271
           |    ||       | |   | |      | ||   ||||     |||||  | |    |
Db    3162 ACGGAGACGGTGTGTTCGAAGGCATAAGGGCATACAACGGACGTGTCTTCAAGTTAAAGG 3221

Qy     272 ACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCA 331
           | ||| | ||| || |    |||  |||         |     |||||        ||  |
Db    3222 AGCACGTTGACAGGCTTTATGACTCGGCAAAAGCAATCGCAATGGACATACCCGTCACAA 3281

Qy     332 AGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGT 391
             || ||| || |||| || | |    ||  | |   ||      | ||    || | |
Db    3282 GAGAGGAAATGACCGAAATAATCCTTGAGGCCCTCAGGAAAAACAACCTCAAAGATGCTT 3341

Qy     392 TCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCA 451
           | |   ||  | |||| |  | |||  |||  | | |   |   ||          || |
Db    3342 ACATCCGTCCCATCGTCTCAAGAGGAATCGGCGACCTCGGGCTTGACCCCCGTAAGTGCG 3401

Qy     452 AGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCG 511
           |    |   |          ||  ||  ||  | |    ||||    | |  ||  |
Db    3402 GGAAACCAAGTG---------TCATCGTTATTGCCCAGGGCTGGGGGGCCATGTACGGAG 3452

Qy     512 AGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACA 571
           | |   |  |          |    ||      | ||||| |    | ||  |
Db    3453 ACCTTTACGAAGTAGGGCTTACCGGAGTTAGCGTCTGTGTCCGGAGAAATGCCCCGGATG 3512

Qy     572 CCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAG 631
           || |   ||||| |||||||  |   | |     |   || |   || |   | ||||
Db    3513 CCCTGTCCCCGAACATCAAGTCCCTGAACTATCTCAACAACATCCTCGCAAAGATAGAAG 3572

Qy     632 CCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAG 691
           |||| ||    || |    |   || ||  | || |||    ||   |  ||   ||| |
Db    3573 CCAACGAGAAGGGAGGAGACGAAGCTATTTTCCTTGACCATAACGGGTTTGTCTGCGAGG 3632

Qy     692 GTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAG---CTGGCCTCCCCGTCCCGGA 748
           |  ||||   |||| | | | ||||||||||| || |   ||| || ||    |   |
Db    3633 GCTCGGGAGACAACATCTTCGTCGTCAAGGACGACAGGGTCCTGACCCCCTTCACAATAA 3692

Qy     749 ACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCG 808
           |   ||  || || ||  |    || |   |||||||    ||| | |||||   |
Db    3693 GCAACCTGAAAGGTATAACAAGGGCCACAGCCATCGAACTTCTGGACGAAATGGGATACA 3752

Qy     809 AAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGG 868
           | |  | |       || ||  |     ||| || |||   || || ||| |  | |
Db    3753 AGGTTATCGAATCAAACCTCGGCCTGTTTGACCTTTACACAGCAGATGAGATCTTTGTCA 3812

Qy     869 TCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACG 928
             || || | ||     | || | | |   || || ||    | | | || | |
Db    3813 CAGGAACTGCAGCCGAATCGGCCCCCGTGACCAGACTTGACGGAAGAGTCATAGGTACAG 3872
```

```
Qy    929 GCGAGCCCGGTCCACTGACGGTGGCCAT 956
          || | ||||| || || ||  ||   ||
Db   3873 GCAAACCCGGGCCTCTCACAATGAAAAT 3900


RESULT 13
PAU63816
LOCUS       PAU63816                7991 bp    DNA     linear   BCT 28-AUG-1998
DEFINITION  Pseudomonas aeruginosa glnE gene, partial cds; ilvE,
            ADP-heptose:LPS heptosyltransferase homolog (waaF),
            lipopolysaccharide heptosyltransferase I homolog (waaC),
            glucosyltransferase I homolog (waaG), RfaP protein (waaP), and
            unknown protein (waaX) genes, complete cds; and inaA gene, partial
            cds.
ACCESSION   U63816
VERSION     U63816.1  GI:3482878
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 7991)
  AUTHORS   Franklund,C.V., Coyne,M.J. and Goldberg,J.B.
  TITLE     Clustering of the Lipopolysaccharide Core Genes, waaF, waaC, waaG,
            and waaP, in Pseudomonas aeruginosa
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7991)
  AUTHORS   Coyne,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1996) Channing Laboratory, Harvard Medical
            School/Brigham & Women's Hospital, 180 Longwood Avenue, Boston, MA
            02115, USA
REFERENCE   3  (bases 1 to 7991)
  AUTHORS   Coyne,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1998) Channing Laboratory, Harvard Medical
            School/Brigham & Women's Hospital, 180 Longwood Avenue, Boston, MA
            02115, USA
  REMARK    Sequence update by submitter
COMMENT     On Aug 28, 1998 this sequence version replaced gi:2226274.
FEATURES             Location/Qualifiers
     source          1. .7991
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAK"
                     /serotype="06"
                     /db_xref="taxon:287"
                     /note="Fisher immunotype IT1; rfa operon"
     gene            <1. .1522
                     /gene="glnE"
     CDS             <1. .1522
                     /gene="glnE"
                     /note="similar to glnE from a variety of prokaryotic
                     cells"
                     /codon_start=2
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC33171.1"
                     /db_xref="GI:3482881"
                     /translation="IPLWEEALDEESASRQLADAGFVDAEAAWKRLSDLRHGPQVRAM
                     QRLGRERLDAFVPRLLAMTVENPQPDLVLERVLPLVEAVARRSAYLVLLTENPGALER
                     LLTLCAASPMVAEQIARFPILLDELLNEGRLFRPPQAAELAAELRERLMRIPEDDLEQ
                     QMETLRHFKLAHGLRVAASEIAGTLPLMKVSDYLTWLAEAILVEVLELAWRQLVQRHG
                     RPLRADGTPCDPDFVIVGYGKVGGLEFGHGSDLDLVFIHDGDPQCETDGGKSIDGAQF
                     FTRLGQKIIHFLTAQTPSGTLYEVDMRLRPSGAAGLLVSSLGAFQRYQEQEAWTWEHQ
                     ALVRARVLAGCRRVQASFEAVRAEVLARPRDLDALRTEVSEMRAKMRDNLGTRATAAG
                     TASNAFEATAAFDLKHDAGGIVDIEFMVQYAVLAWSGEHPALLEFTDNIRILEGLERA
                     GLIASEDVRLLQEAYKAYRAAAHRLALQKEAGVVSGEHFQTERREVIRIWRELRLG"
     gene            1578. .2501
                     /gene="ilvE"
     CDS             1578. .2501
                     /gene="ilvE"
                     /note="similar to ilvE from Escherichia coli and other
                     organisms"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC33172.1"
                     /db_xref="GI:3482882"
                     /translation="MSMADRDGVIWYDGELVQWRDATTHVLTHTLHYGMGVFEGVRAY
                     DTPQGTAIFRLQAHTDRLFDSAHIMNMQIPYSRDEINEATAPPCARTTWKAPISARWC
                     STGSEGMGLRASGLKVHVIIASWSWGAYMGEEALQQGIKVRTSSFTRHHVNISMTRAK
                     SNGAYINSMLALQEAISGGADEAMMLDPEGYVAEGSGENIFIIKDGVIYTPEVTACLN
                     GITRNTILTLAAEHGFKLVEKRITRDEVYIADEAFFTGTAAEVTPIREVDGRKIGAGR
                     RGPVTEKLQKAYFDLVSGKTEAHAEWRTLVK"
     gene            2561. .3598
                     /gene="waaF"
     CDS             2561. .3598
                     /gene="waaF"
                     /note="similar to Escherichia coli RfaF protein encoded by
                     GenBank Accession Number U00039"
                     /codon_start=1
                     /transl_table=11
                     /product="ADP-heptose:LPS heptosyltransferase homolog"
                     /protein_id="AAC33166.1"
                     /db_xref="GI:3482879"
                     /translation="MRILIVGPSWVGDMVMAQTLFQCLRQRHPECVIDVLAPEWSRPI
                     LERMPEVRQALSFPLGHGVMDVATRRRIGRGLRGQYEQAILLPNSLKSALVPWFAGIP
                     KRTGWRGEMRYGLLNDIRKLDKQRYPLMIERFMALAFEPGVELPKPYPQPRLRIDDGS
                     RQAALDKFALSLDRPVLALCPGAEFGEAKRWPAEHYAAVAEAKIRAGWQVWLFGSKND
                     HPGGEEIRQRLIPGLREESFNLAGETSLAEAIDLMSCAGAVVSNDSGLMHVAAALDRP
                     LVGVYGSTSPQFTPPLADRVEIVRLGLECSPCFERTCRFGHYNCLRELPPGLVLQALE
                     RLVGDPAEVAG"
     gene            3595. .4662
                     /gene="waaC"
     CDS             3595. .4662
                     /gene="waaC"
                     /note="similar to  Escherichia coli RfaC protein encoded
                     by GenBank Accession Number U00039"
                     /codon_start=1
                     /transl_table=11
                     /product="lipopolysaccharide heptosyltransferase I
                     homolog"
                     /protein_id="AAC33167.1"
                     /db_xref="GI:2226276"
                     /translation="MRVLLVKTSSLGDVIHTLPALTDAARAIPGIQFDWVVEEGFAEI
                     PAWHPAVARVIPVAIRRWRKNLWQTLRNGEWRRFKQRLKEVDYDLVIDAQGLLKSAWL
                     TRYVGKTPVAGLDRDSAREPLASRFYRRAYPVAWGQHAVERTRQLFAQALDYPLPESV
                     GDYGLDREQLADADPGAPYLVFLHGTTWDTKHWPEAYWRELAERMCERGWSVRLPWGS
                     AAERERAGRLAAGLENAAVLPRLSLAGMAKVLAGARACVAVDTGLGHLAAALDVPTLS
                     LFGPTNPGFTGAYGRSQVHLGSDFPCAPCLKKTCTYQPTEEDRKLFDLKREQPLCFTR
                     LNPQRVATQLEAMLLAPETLR"
     gene            4659. .5780
                     /gene="waaG"
     CDS             4659. .5780
                     /gene="waaG"
                     /note="similar to  Escherichia coli RfaG encoded by
                     GenBank Accession Number U00039"
                     /codon_start=1
                     /transl_table=11
                     /product="glucosyltransferase I homolog"
                     /protein_id="AAC33168.1"
                     /db_xref="GI:2226277"
                     /translation="MTLAFILYKYFPFGGLQRDFMRIALECQRRGHDIRVYTLIWEGD
                     VPDGFEVLVAPVRSIFNHRRNEKFTAWVRADLDRRPVQRVIGFNKMPGLDVYYAADAC
                     FEEKAQTLRNPLYRQWGRYRHFAGYERAVFDPASKTEILMISEVQQPLFVKHYGTQAE
                     RFHLLPPGISQDRRAPANAADVRAEFRREFGLEEDDLLLVQIGSGFKTKGLDRSLKAL
                     AALPKALRRRTRLIAIGQDDPKPFLLQIAALGLNDQVQILKGRSDIPRFLLGADLLIH
                     PAYNENTGTVLLEALVSGLPVLVTDVCGYAYYIAEADAGRVLPSPFEQDSLNRLLAEM
                     LEDAPARAAWSRNGLAYADHADLYSMPQRAADLILGEAS"
     gene            5777. .6583
                     /gene="waaP"
     CDS             5777. .6583
                     /gene="waaP"
                     /function="putatively involved in phosphorylation of LPS
                     core heptose"
```

```
                     /note="similar to Escherichia coli RfaP protein encoded by
                     GenBank Accession Number U00039"
                     /codon_start=1
                     /transl_table=11
                     /product="RfaP protein"
                     /protein_id="AAC33169.1"
                     /db_xref="GI:2226278"
                     /translation="MRLVLEEPFKRLWSGRDPFEAVEALQGKVYRELEGRRTLRTEVD
                     GRGYFVKIHRGIGWGEIAKNLLTAKLPVLGARQEWQAIRRLHEAGVTTMTAVAYGERG
                     SDPARQHSFIVTEELAPTVDLEVFSQDWRERPPPPRLKRALVEAVARMVGDMHRAGVN
                     HRDCYICHFLLHTDKPVSADDFRLSVIDLHRAQTRDATPKRWRNKDLAALYFSALDIG
                     LTRRDKLRFLRTYFRRPLREILRDEAGLLAWMERKAEKLYERKQRYGDLL"
     gene            6583. .7317
                     /gene="waaX"
     CDS             6583. .7317
                     /gene="waaX"
                     /note="similar to waaP"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein"
                     /protein_id="AAC33170.1"
                     /db_xref="GI:3482880"
                     /translation="MAGWTLEPDYRHLAADFGSLDSVFALEGERLTRDPLSEVIRIER
                     DGVRYYVKRYRSAGKGLRRYLARPRIKAEWQNLKRFEKWGIPTAEVVAWGLERKAGAF
                     QRGAMITRELPGTEDLSVLACNRDPRLKDPRWVDGVSRQIAKATRLMHDQHFTHNDLK
                     WRNLLVDAQPLVYFIDCPNGAFWWSFMLRYRITKDLACLDKVAKYHLSRTQRLRFYLQ
                     YRRRRHLNESDKKRIRQVVAFFEGRE"
     gene            7314. .>7991
                     /gene="inaA"
     CDS             7314. .>7991
                     /gene="inaA"
                     /note="similar to inaA from Escherichia coli protein
                     encoded by GenBank Accession Number AE000313"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC33173.1"
                     /db_xref="GI:3482883"
                     /translation="MSVFIAAQDHDLLQRNRLDSFDALWALQLDAVDEPNTERGGWSS
                     VYRLDLGDSAYYLKRQSNHLTRSLRHPLGEPTFAREFRNIERYRALGIPALQAAFFAQ
                     QAVGGERQAILLTRALDGWQDLDTWLASWHDLAEERRRAILRAVGRLARQLHAAGQMH
                     GCFYPKHVFLREGADGVFEACLIDLEKTRPIWFGRRDQVKDLEPMLRRAPDWSEADVH
                     EFISAYVD"
BASE COUNT     1330 a   2690 c   2667 g   1304 t
ORIGIN

  Query Match             8.2%;  Score 83.2;  DB 1;  Length 7991;
  Best Local Similarity   48.1%;  Pred. No. 0.011;
  Matches  337;  Conservative    0;  Mismatches  348;  Indels   15;  Gaps    3;

Qy     256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
           |||||||| ||    |  ||    |||||| || |||||   |||  ||   ||  | ||
Db    1731 ATCTTCCGCCTGCAGGCGCATACCGACCGGCTGTTCGACTCCGCGCACATCATGAACATG 1790

Qy     316 GACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCG 375
             |   |  |||||||         |  | ||||||| | ||  |  || | ||  ||
Db    1791 CAGATCCCGTACAGC---------CGCGACGAGATCAACGAGGCGACCGCGCCGCCGTGC 1841

Qy     376 CAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGC 435
             |         | |   | |  | |  ||| |    || | |              || | |||
Db    1842 GCGAGAACAACCTGGAAAGCGCCTATATCCGCCCGATGGTGTTCTACGGGAAGCGAAGGC 1901

Qy     436 GAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGG 495
             |       |   |||     |     |           |||  ||||  |  |  |  | |
Db    1902 ATGGGCCTGCGCGCCAGCGGCCTGAAGGTCCATGTGATCATCGCCTCCTGGAGCTGGGGC 1961

Qy     496 GCCTTCCCGCCCGCCGA---GCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
           |||| |   |  ||  ||    | |      |||| ||  | |  |         |     |
Db    1962 GCCTACATGGGCGAGGAAGCCCTGCAGCAAGGCATCAAGGTGCGCACCAGTTCCTTCACC 2021

Qy     553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
           |||| |  || |  || | |  |   ||   |||     |||  |   |       |
Db    2022 CGCCACCACGTCAACATCTCGATGACCCGTGCCAAGTCCAACGGCGCCTACATCAACTCG 2081

Qy     613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
           |     |||      ||||| |    | | || ||   |   || ||  |||||||  ||
Db    2082 ATGCTGGCCCTCCAGGAAGCGATCTCCGGCGGCGCCGACGAGGCCATGATGCTCGATCCG 2141

Qy     673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
           ||   || ||||||||||||  | |||   ||| | | |||| |||||||  |||  | |
Db    2142 GAAGGCTACGTGGCCGAAGGCTCCGGCGAGAACATCTTCATCATCAAGGATGGCGTGATC 2201

Qy     733 GCCTCCCCGTCCCGGAACGCGTTGCCG---GGCATCACCCGTAAGACGGTGTTCGAACTG 789
             | |||||      | |||  |  |   |||||||| ||||| ||  |  |    |||
Db    2202 TACACCCCGGAAGTCACCGCCTGCCTGAACGGCATCACTCGTAACACTATCCTGACCCTG 2261

Qy     790 GCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGAC 849
           |||| | |    ||  |  ||     |  |     | ||||| ||   ||  | |||  |
Db    2262 GCCGCCGAACACGGTTTTAAACTGGTCGAGAAGCGCATCACCCGCGACGAGGTGTACATC 2321

Qy     850 GCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGAT 909
           |||||||||    |          ||| || |  |  ||||| ||||||  |    ||||  .
Db    2322 GCCGACGAGGCCTTCTTCACTGGCACTGCCGCGGAAGTCACGCCGATCCGCGAAGTGGAC 2381

Qy     910 GGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGG 949
           ||         | |||  ||||   |  || ||  | || |
Db    2382 GGTCGCAAGATCGGCGCCGGCCGCCGTGGCCCGGTCACCG 2421


RESULT 14
AE004996
LOCUS       AE004996               10196 bp    DNA     linear   BCT 12-FEB-2001
DEFINITION  Halobacterium sp. NRC-1 section 27 of 170 of the complete genome.
ACCESSION   AE004996 AE004437
VERSION     AE004996.1  GI:10580000
KEYWORDS    .
SOURCE      Halobacterium sp. NRC-1.
  ORGANISM  Halobacterium sp. NRC-1
            Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE   1  (bases 1 to 10196)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Genome sequence of Halobacterium species NRC-1
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
  MEDLINE   20504483
   PUBMED   11016950
REFERENCE   2  (bases 1 to 10196)
  AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES             Location/Qualifiers
     source          1. .10196
                     /organism="Halobacterium sp. NRC-1"
                     /strain="NRC-1"
                     /db_xref="taxon:64091"
     gene            257. .376
                     /gene="trpG2"
                     /note="VNG0386G"
     CDS             257. .376
```

```
                     /gene="trpG2"
                     /note="TrpG2"
                     /codon_start=1
                     /transl_table=11
                     /product="anthranilate synthase beta chain"
                     /protein_id="AAG18943.1"
                     /db_xref="GI:10580001"
                     /translation="MGVRHRDAPHVGVQFHPESILTPVGSELITNFLSLCDTT"
     gene            361. .1218
                     /gene="ilvE1"
                     /note="VNG0387G"
     CDS             361. .1218
                     /gene="ilvE1"
                     /note="IlvE1"
                     /codon_start=1
                     /transl_table=11
                     /product="branched-chain amino acid aminotransferase"
                     /protein_id="AAG18944.1"
                     /db_xref="GI:10580002"
                     /translation="MRYHVNGALVDAADATVSVRDRGFQYGDAAFETMRAYGGQTFAW
                     TAHYRRLQATCDSLGIDHGLSEAALRGRVHETLAANDLADAYVRLSITRGVQDGRLTP
                     RDTTDPTVVVVVDALPRGGVDGDPVWSMPATVETTTVTQHRDTTIPSTAKTHNYLPGV
                     LARRDIAADADEALFVTPDGHLTEGAASNLFFVADGTLYTPSGDLDVLPGITRWAVAE
                     LAAEHGLTVHTGAYTPSRLYDADEAFLTNTTWEVRPIARVDDATYDTCRIGAALASAY
                     NAEIERRHY"
     gene            1264. .1734
                     /gene="VNG0388C"
     CDS             1264. .1734
                     /gene="VNG0388C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0388c"
                     /protein_id="AAG18945.1"
                     /db_xref="GI:10580003"
                     /translation="MVDGDTDGGDAAEPAPGDGAQIASVTDLQAGDSFRFTVRDRDSG
                     DEQEAILVALSDGSITGWLNYCMHWTDVTLDSGSGAVRDGELVCRKHAATFDPTSGVC
                     THGPCEGATLTAVDVTVQDGTVSLTDPDYEFVMQGVDPGDPADLSTSPGSRIGF"
     gene            complement(1731. .2021)
                     /gene="VNG0389C"
     CDS             complement(1731. .2021)
                     /gene="VNG0389C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0389c"
                     /protein_id="AAG18946.1"
                     /db_xref="GI:10580004"
                     /translation="MDDRTTRERVRAALREQPATPSEVAAAFEVARGTALTHVRHIAE
                     SLDSADEELLVRPPACRDCGFEDFDDPVNVPSRCPECKSESIAEPAFVVESA"
     gene            2041. .3069
                     /gene="graD1"
                     /note="VNG0390G"
     CDS             2041. .3069
                     /gene="graD1"
                     /note="GraD1"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose-1-phosphate thymidylyltransferase"
                     /protein_id="AAG18947.1"
                     /db_xref="GI:10580005"
                     /translation="MKKDLPDRVDLVGMKAVVLAGGYATRLWPITKDRPKMFLPVGDS
                     TVIDRIFADLESDDRIDTVYVSTNERFADEFRSHLAERSFEKPQLSVEDTTDENEKFG
                     VMGALAQLVDRESITDDLVVIAGDNLISFDVSAFVDFFAADDADNPAIAAYDVGSYER
                     AQSYGLVELHGDEVINFQEKPDDPNSTLVSIACYAYPESVVEDIDTYLAADGNPDEPG
                     WFVEWLQDRQSVRAFTFEGAWFDIGTPESYLDAVAWTLDGDSIVADDATVSNTTIGEN
                     VHVLPGAELTNSTVSNSVIFPDATLRDCDVRNSIIDEDTVLDSIDFSGALIGAHTTIE
                     NGHDADGR"
     gene            complement(3070. .3792)
                     /gene="VNG0391C"
     CDS             complement(3070. .3792)
                     /gene="VNG0391C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0391c"
                     /protein_id="AAG18948.1"
                     /db_xref="GI:10580006"
                     /translation="MTDGQWVGLFSGGKDSSWALYRALEAGLDVGRLVTVHPESDSYM
                     YHVPATRLAELAAESIGIPVVNVTPDAFETADVADAGAQGDTELEPLEDAVTDVADTL
                     DDGLAGVIAGAVESDFQTSRMEGMCDRLGCELYAPLWQQPPRELADAMLTAGFEIRII
                     QVAAYGLDKSWLGRRLDRDALAALAALNDEYGVHMLGEGGEFETLVTDAPHMDRPIRL
                     DYETVWAGDRGHIEITEACLGP"
     gene            complement(3839. .4987)
                     /gene="VNG0393C"
     CDS             complement(3839. .4987)
                     /gene="VNG0393C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0393c"
                     /protein_id="AAG18949.1"
                     /db_xref="GI:10580007"
                     /translation="MRTLVVCVDRAGDISRQTGLRTPIAGWEAVQSLVTDMGVADPED
                     ASVNCLLEALRVTRDLRDGADDAVVAAVSATGDGVTADRSVAQQFDSLVDRHDPDSCV
                     VVVDSAADERVVPIVESRIQVDAVDRVVVRQARDLESTYYLLKQFLADEELRQTVLVP
                     VGIVLLVFPGLMMVTGSLAVAAASITTVIGLFLLYKGLGVDEYVAALPGQARDALYSG
                     RVSIVTYVVAGGLGLVGVFAGALGLTNQAAGLSRLVAAMTFTYYSIPWVTLGALTASG
                     GRLFDEYIRNDDVRTSFVNLPFGVLAVGLVVRGFAVYFMERAGIVAPLTAPSAALGAL
                     SVRGFQVSPEQRLAVFVVLGVLVSVAGIRVATYVSGVTIRDVERSAEQ"
     gene            complement(5061. .5963)
                     /gene="VNG0394C"
     CDS             complement(5061. .5963)
                     /gene="VNG0394C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0394c"
                     /protein_id="AAG18950.1"
                     /db_xref="GI:10580008"
                     /translation="MAEEETTIEVSTADELITDDELQNKSKGQLIKNAGQFRDRRNEL
                     NQLASSRASERDDLNAKTREKVDEAQEHREKRDELNERVQEHKEIRNDLNADANELFD
                     EVEQRKEDLELDEGKDLEQLKEEIEELEFKQQTEVLSTEDERELIEKIEGKREEYQDR
                     EEKLDQSGDLEGLVEEAESVRAEASKHHEKVTELADKAQEHHNQMIEAYREADDIRDE
                     ADEMHEDFVEAQEAADAHHEAFVEVQKRLRELDKQEEEEEKDKRAQEQEEAREEAEEI
                     YERFKEGETLDTEDLRKLQKSGHL"
     gene            6224. .6637
                     /gene="VNG0396C"
     CDS             6224. .6637
                     /gene="VNG0396C"
                     /note="conserved hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0396c"
                     /protein_id="AAG18951.1"
                     /db_xref="GI:10580009"
                     /translation="MDVSVSATGHDNVTARHDSTLELTTDDFLTPAGDCILGIDADCS
                     PAAFSDAFVAACQDAAATITLVLDTPQHRDEITGSGHPDLTFTNDRSMVGRTSTYVDD
                     RTVFVDADKAAADVDRELVAALADGAPLTATLRVD"
     gene            6695. .7492
                     /gene="nthA1"
                     /note="VNG0398G"
     CDS             6695. .7492
                     /gene="nthA1"
                     /note="NthA1"
                     /codon_start=1
                     /transl_table=11
                     /product="endonuclease III"
                     /protein_id="AAG18952.1"
                     /db_xref="GI:10580010"
                     /translation="MADEPAENIAGGDDSRPTTFASGADPTQAALVVDRLGELYWQKT
                     YGGRAAFECLVRTVLSQNTSDKASQPAHDALMQRYGAAADDLAAALADADQDVLADTI
                     SMAGLYNQKSATIIRIADRVCEEYGGADGFDAFVREGDPDAVRAALLDMTGVGPKTAD
                     CVLLFAGGAPGVFPVDTHVHRIARRMGIAPAAADHEAVRAALEAAVPDAACGFGHTAM
                     IQFGREYCTARDPACLDDPAACPLAEHCDRVGVDAATGAVVDPADAD"
     gene            complement(7494. .7961)
```

```
                     /gene="VNG0399H"
     CDS             complement(7494. .7961)
                     /gene="VNG0399H"
                     /note="hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Vng0399h"
                     /protein_id="AAG18953.1"
                     /db_xref="GI:10580011"

  Query Match             8.1%;  Score 82.6;  DB 1;  Length 10196;
  Best Local Similarity   46.7%;  Pred. No. 0.013;
  Matches  370;  Conservative    0;  Mismatches  414;  Indels    9;  Gaps    3;

Qy    152 AGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTC 211
Db    377 ACGGCGCGCTCGTGGACGCCGCCGACGCAACCGTCAGCGTCCGCGACCGCGGCTTCCAGT 436

Qy    212 ATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCG 271
Db    437 ACGGCGACGCAGCGTTCGAAACCATGCGCGCGTACGGCGGCCAGACGTTCGCCTGGACCG 496

Qy    272 ACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCA 331
Db    497 CCCACTACCGCCGCCTGCAGGCCACCTGTGACTCGCTGGGCATCGACCACGGCCTCTCGG 556

Qy    332 AGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGT 391
Db    557 AGGCGGCGCTGCGGGGGCGCGTCCACGAGACGCTCGCCGCCAACGACCTCGCGGACGCCT 616

Qy    392 TCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGA---GAAGGACCTGT 448
Db    617 ACGTCCGGCTGTCGATCACCCGGGGCGTCCAGGACGGCCGGCTCACCCCGCGGGACACGA 676

Qy    449 CCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTG---GGCCTTCCCGC 505
Db    677 CCGATCCGACCGTCGTCGTGGTGGTCGACGCGCTCCCCCGCGGCGGCGTCGACGGCGACC 736

Qy    506 CCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCC 565
Db    737 CGGTCTGGTCGATGCCCGCCACCGTCGAAACCACGACGGTCACCCAACACCGCGACACCA 796

Qy    566 GCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTT 625
Db    797 CCATCCCCAGCACCGCGAAAACCCACAACTACCTGCCCGGCGTGCTCGCCCGCCGCGACA 856

Qy    626 TCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGG 685
Db    857 TCGCCGCCGACGCCGACGAAGCGCTGTTCG---TCACGCCCGACGGCCACCTCACGGAGG 913

Qy    686 CCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCC 745
Db    914 GCGCTGCGAGCAACCTCTTTTTCGTCGCCGACGGCACCCTCTACACGCCCAGCGGCGATC 973

Qy    746 GGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCA 805
Db    974 TCGACGTGTTGCCCGGGATCACGCGGTGGGCGGTCGCCGAACTCGCCGCCGAGCACGGGC 1033

Qy    806 TCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGG 865
Db   1034 TCACGGTGCACACGGGCGCGTACACGCCGTCTCGGCTCTACGACGCCGACGAGGCGTTCC 1093

Qy    866 CGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCA 925
Db   1094 TCACGAACACCACCTGGGAGGTGCGCCCCATCGCGCGTGTCGACGACGCCACGTACGACA 1153

Qy    926 ACGGCGAGCCCGG 938
Db   1154 CGTGCCGGATCGG 1166


RESULT 15
AE013330
```

```
LOCUS       AE013330               10029 bp    DNA     linear   BCT 17-MAY-2002
DEFINITION  Methanosarcina mazei strain Goe1, section 112 of 379 of the
            complete genome.
ACCESSION   AE013330 AE008384
VERSION     AE013330.1  GI:20905468
KEYWORDS    .
SOURCE      Methanosarcina mazei Goe1.
  ORGANISM  Methanosarcina mazei Goe1
            Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
            Methanosarcinaceae; Methanosarcina.
REFERENCE   1  (bases 1 to 10029)
  AUTHORS   Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
            Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,
            Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,
            Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
            Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
  TITLE     The genome of Methanosarcina mazei: Evidence for lateral gene
            transfer between Bacteria and Archaea
  JOURNAL   J. Mol. Microbiol. Biotechnol. (2002) In press
REFERENCE   2  (bases 1 to 10029)
  AUTHORS   Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
            Martinez-Arias,R., Henne,A., Wiezer,A., Baeumer,S., Jacobi,C.,
            Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,
            Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
            Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute
            for Microbiology and Genetics, Grisebachstrasse 8, Goettingen
            37077, Germany
FEATURES             Location/Qualifiers
     source          1. .10029
                     /organism="Methanosarcina mazei Goe1"
                     /strain="Goe1"
                     /db_xref="taxon:192952"
     gene            complement(149. .628)
                     /gene="MM1035"
     CDS             complement(149. .628)
                     /gene="MM1035"
                     /EC_number="2.3.1.1"
                     /codon_start=1
                     /transl_table=11
                     /product="Amino-acid acetyltransferase"
                     /protein_id="AAM30731.1"
                     /db_xref="GI:20905469"
                     /translation="MQLQRTQKLIRKARISDVVEMKQIINTYSREELMLARSLSEMYE
                     NIRDYYVCEIEGEIVGCCALHVVWEDLAEILALAVKPDCARKGVGTKLVSACLEDAKS
                     LGMKEVFALTYVPGFFETLGFNIVDKNSLPRKIWSGCIRCPKFPDCNEIAVLKSIKE".
     tRNA            complement(841. .915)
                     /product="tRNA-Arg"
     gene            1356. .1919
                     /gene="MM1036"
     CDS             1356. .1919
                     /gene="MM1036"
                     /codon_start=1
                     /transl_table=11
                     /product="BioY protein"
                     /protein_id="AAM30732.1"
                     /db_xref="GI:20905470"
                     /translation="MHIRHDNSYPPHTSELRKMVFASLFAALTAVGAYIQIPIPFSPV
                     PVTLQVFFVLLAGSMLKSKWGSLSMVVYTLLGVAGLPVFSGGSSGLGVLLGPTGGYIF
                     GFIVAAYFTGKLAEKAERAGKSGLAVNALNMSAGILTIYALGVTQLMLVAEIGLGTAL
                     ALGAIPFIPGEIVKTAAAAYIATTYKL"
     gene            2005. .2829
                     /gene="cbiO"
                     /note="MM1037"
     CDS             2005. .2829
                     /gene="cbiO"
                     /codon_start=1
                     /transl_table=11
                     /product="Cobalt transporter, ATP-binding protein"
                     /protein_id="AAM30733.1"
                     /db_xref="GI:20905471"
                     /translation="MIRLENVSYCYPDGTPALTNINLTIRKGEYLGIIGRNGSGKSTL
```

```
                ALHLNGLRRPQKGKVIVKGIDAGDFSKLQEVRKIVGIVFQNPETQFIGRTVEEDLAFG
                PENLCLPPTEIRKRVDRALAEIKLEKYRHHSPKSLSGGQGQCVALAGVLTMEPECLVF
                DEVTSMLDPDSGETVLRRIKKLHEKGKTIVYTTHNLEELHVADRILVMDRGRIVLEGE
                PENVLSDPSLKDLGLTPPSLIELAEQLKKHEIEISWENTSSPSSFAEELCRLFLKT"
gene            2763. .3668
                /gene="cbiO"
                /note="MM1038"
CDS             2763. .3668
                /gene="cbiO"
                /codon_start=1
                /transl_table=11
                /product="Cobalt transporter, ATP-binding protein"
                /protein_id="AAM30734.1"
                /db_xref="GI:20905472"
                /translation="MGKYLFPFQLCGGTMPIILENISVFYSRNTPLEISALKDVNLRI
                EKGEFVGILGEKGAGKSTLIKLFNGLLRPDSGRITVNGLDPSSKKVKSRVGMVFQQPA
                DQLFCRTVYEEIAFGPLNFGYSRKETEERVFEALEAASLERSMLSRDPLSLSGGEMQR
                VALAGALALRPDYLVLDEPITGLDPAGKKEILETLKKIKGQGTTIITVTHNLKGFFPL
                LERIVLIKEGRISFQGSRKEYMETENVPLPPVASMMRELHSRGIQVNPAVFTVEEALK
                EILRVKSMIEKEKAEKERAEKQTKK"
gene            3739. .4512
                /gene="cbiQ"
                /note="MM1039"
CDS             3739. .4512
                /gene="cbiQ"
                /codon_start=1
                /transl_table=11
                /product="Cobalt transport protein"
                /protein_id="AAM30735.1"
                /db_xref="GI:20905473"
                /translation="MQEPVFSYVPGESFLHKLDPRTKLAAVMLLGILAFRAESFLEIG
                TLFTFFIALASFSRLPLKVFFRAVRPMALFIIFIFLAQFFFTEGQELASFWVLRPTLE
                GLYNGIRISARFILLLLFAALMTASTDPSAITCGIERMLRPLPLRWLGVNSYELATMM
                NLSIAFLPLLFDRVERTKAAQASRGMYFGKNPLRSVPALTIPLIRGVVRDAEELSLAM
                ESRGYQGTCRTSIHELLMRKRDRLSLLVMGLFAIIMLVL"
gene            4620. .4967
                /gene="MM1040"
CDS             4620. .4967
                /gene="MM1040"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /protein_id="AAM30736.1"
                /db_xref="GI:20905474"
                /translation="MQNKCKRVSTEQIKTLMQKVPDAEIITRMSAIFQALQSDTRLKI
                LFLLRHKEMCVCELEEELEVTQSAVSHGLRTLRQLDLVRVRREGKFTVYYIADEHVRT
                LIEMCLEHVEEKA"
gene            4964. .6331
                /gene="MM1041"
CDS             4964. .6331
                /gene="MM1041"
                /codon_start=1
                /transl_table=11
                /product="Transporter"
                /protein_id="AAM30737.1"
                /db_xref="GI:20905475"
                /translation="MSTDIFSSIPELISGILISSWEIFVEAAPYLIFGFAIAGILNVL
                VPDQKIVEYLGTSAGKIRSVINAALAGLPLPLCSCGVIPAAMSIRKRGANRGATLSFL
                ISTPQTGMDSISITYALLDPLMAVFRPLATLATAILAGIADNLLIGEDPEKSGKSENE
                KAGKKAEILAVSTLVGVSVSGDKCSPVSCNCGTHEATETGKEKHPDKNTVPLKSGTLE
                LKSETGRETSHLKITPAIHASENNTPSCGCGHCGEGKREQFSDKRVKKSIKEDILECL
                KYAFIELPGEIAKWMLIGILLAGIISYAIPETLVQEYLGGGFSSMLVMLVIGIPLYIC
                ATASTPLAASLVAKGMSPGTAFVFLLAGPATNAATITMVARFLGKRSAVLYVGVISLC
                AIGAGILLDWLYFKLGISVTATLGSAEEMLPESIKLGFAVLLLPLMLYGIFRRNHGRG
                CTECH"
gene            6424. .6930
                /gene="MM1042"
CDS             6424. .6930
                /gene="MM1042"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator"
                /protein_id="AAM30738.1"
                /db_xref="GI:20905476"
                /translation="MQENSEFQEKKEPEESDFKSPGELETGYTEKSWKKPEKTMEEKG
                TTMDSEYQEREEPASSGFKAPGEHEAGYTERSWKKPEGTMEEDIGSGGETREEESTGY
                TKKRLTRSKSDRMVFGVCGGLGKYFGVDPTLIRLGFAFFILLSGIGPGIVIYIIMAII
                MPQENNVK"
gene            7277. .7771
                /gene="MM1043"
CDS             7277. .7771
                /gene="MM1043"
                /codon_start=1
                /transl_table=11
                /product="Transposase"
                /protein_id="AAM30739.1"
                /db_xref="GI:20905477"
                /translation="MIKYTVTLTEDERRSLCELASKGKHNSQQILNALILLNCDKSEL
                NVSHSTNEEISRVLNISMKKIDRVKKRFVEEGLEVALNGKESERIYNKKVDGDLEAHL
                VALSCSQPPEGFARWSLRLLADKAVELGYFEEISHETVRRTLKKRNQTLAKETMGNSS
                RTKQ"
gene            7785. .8408
                /gene="MM1044"
CDS             7785. .8408
                /gene="MM1044"
                /codon_start=1
                /transl_table=11
                /product="Transposase"
                /protein_id="AAM30740.1"
                /db_xref="GI:20905478"
                /translation="MEMVLDVYKRPFDPRNPVVCMDESPKQLIAETRIPIPCSTGKPE
                KYDYEYKRNGMCNIFLACEPLTGKRMVKITERKTKRDWAYFLEEIEVQHENADKITLV
                MDNLNTHIPGSLYETFPPPKAKALWDRFEFVYTPKHGSWLNMAEIELNVLTGQCLKRR
                MDNIEFVRKEVLAWQNYRNNKNSKVKWQFTTDDARIKLSRLYPTIED"
tRNA            complement(8581. .8655)
                /product="tRNA-Arg"
gene            8996. .9880
                /gene="MM1045"
CDS             8996. .9880
                /gene="MM1045"
                /EC_number="2.6.1.42"
                /codon_start=1
                /transl_table=11
                /product="Branched-chain amino acid aminotransferase"
                /protein_id="AAM30741.1"
                /db_xref="GI:20905479"
                /translation="MQMSELLIYLDGKFVPKSEAKVSVYDHGYLYGDGVFEGIRAYNG
                RVFKLKEHVDRLFDSAKAIAMDIPVTREEMTEIILETLRKNNLKDAYIRPIVSRGVGD
                LGLDPRKCEKPCVVVIAQGWGAMYGNLYEIGLTGVSVCVRRNAPDALSPNIKSLNYLN
                NILAKIEANEKGGDEAIFLDQNGFVCEGSGDNIFIVKNDKVFTPHTISNLKGITRATA
                IELLDEMGYKVIEANLGLFDLYTADEIFVTGTAAESAPVTRLDGRIIGNGKPGPLTMK
                MVEAFDKITRSTGTPIYK"
BASE COUNT      3180 a    2040 c    2233 g    2576 t
ORIGIN

  Query Match             7.1%;  Score 72.4;  DB 1;  Length 10029;
  Best Local Similarity   50.2%;  Pred. No. 0.32;
  Matches  206;  Conservative    0;  Mismatches  201;  Indels    3;  Gaps    1;

Qy     550 GTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGAT 609
           |||||| |    | |   |  || |   ||| | |||||||  |    | |      ||
Db    9416 GTCCGCAGGAATGCACCTGATGCCCTGTCCCCTAACATCAAGTCCCTGAACTACCTGAAT 9475

Qy     610 CTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGAC 669
             | |   || |   | ||||| || ||    || |        || ||  | || |||
Db    9476 AATATCCTTGCAAAGATAGAAGCAAATGAGAAAGGCGGAGATGAAGCCATATTCCTTGAC 9535

Qy     670 TCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAG 729
             | ||  ||  |||  ||||||| | ||   |||| | |  |||||||  | |  |||
Db    9536 CAGAACGGCTTTGTGTGCGAAGGTTCTGGAGACAACATCTTTATCGTCAAAAATGATAAG 9595

Qy     730 CTGG---CCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAA 786
           |     |  |||   |    | |   ||    ||||||||  |    || |   | ||
Db    9596 GTCTTTACTCCCCACACAATAAGCAACTTAAAAGGCATCACAAGAGCTACAGCCATTGAG 9655

Qy     787 CTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTAC 846
```

```
          ||    ||  | |||||   | | |  |         || ||  |     ||| ||||||
Db   9656 CTTCTGGATGAAATGGGATACAAGGTTATTGAAGCAAACCTCGGCCTTTTTGACCTCTAC 9715

Qy    847 GACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTG 906
          || |||||  |  | |    |   || || | ||      | ||| ||| |  |||
Db   9716 ACGGCAGACGAAATTTTCGTTACCGGGACTGCAGCCGAATCTGCTCCGGTCACCAGGCTC 9775

Qy    907 GATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCAT 956
          || ||      | | || || ||| | |||||||| || ||  ||   ||
Db   9776 GACGGAAGAATCATCGGTAATGGCAAACCCGGTCCTCTCACAATGAAGAT 9825


Search completed: April 24, 2003, 22:21:04
Job time : 3997 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 19:00:47 ; Search time 279 Seconds
(without alignments)
8233.112 Million cell updates/sec

Title: US-10-067-291-2
Perfect score: 1020
Sequence: 1 atgactgctctttcagacct.....tcgaaacgatcgaatactga 1020

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1020	100.0	1020	21	AAA94032	Mycobacterium auru
2	1018.4	99.8	1020	21	AAA94040	Mycobacterium auru
3	326.6	32.0	993	19	AAV83675	DNA encoding a pro
4	326.6	32.0	993	19	AAV83676	DNA encoding a pro
5	326.2	32.0	978	19	AAV73146	DNA encoding a pro
6	325	31.9	1574	19	AAV73156	PstI fragment of a
7	66.2	6.5	15872	18	AAT68715	Streptomyces venez
8	66.2	6.5	15872	21	AAZ87283	S. venezuelae vep
9	64.2	6.3	58857	21	AAA58471	Nucleotide sequenc
10	59	5.8	1208	15	AAQ64204	snaC gene encoding
11	58.4	5.7	9169	24	ABK91617	Modified HIV prote
12	58.4	5.7	9785	24	ABK91624	Modified HIV prote
13	57.8	5.7	12588	15	AAQ63293	Sequence encoding
14	57.4	5.6	5392	15	AAQ64201	Sequence comprisin
15	57.2	5.6	2271	11	AAQ06844	Amylase gene from
16	57	5.6	1140	19	AAV41733	Codon-optimised RA
17	56.2	5.5	1053	22	AAH44045	Streptomyces sp. C
18	56.2	5.5	1065	22	AAH78255	Nucleotide sequenc
19	56.2	5.5	1065	24	ABL41306	Streptomyces IPP r
20	56.2	5.5	6798	22	AAH44043	Streptomyces sp. C
21	56.2	5.5	6798	22	AAH78258	Nucleotide sequenc
22	56.2	5.5	6798	24	AAD31022	Streptomyces sp. C
23	56.2	5.5	8077	24	AAD31028	Operon F DNA encod
24	55.2	5.4	2821	22	AAI66075	Triticum aestivum
25	55	5.4	2291	9	AAN80309	Entire amylase gen
26	54.6	5.4	1162	21	AAX89451	Rice branched chai
27	53.8	5.3	1294	20	AAZ06824	Streptomyces albid
28	53.8	5.3	1294	22	AAH74537	Nucleotide sequenc
29	53.6	5.3	2634	14	AAQ39093	Streptomyces nodos
30	53.4	5.2	12152	22	AAS08699	M. carbonacea DNA
31	52.8	5.2	1509	24	ABL39958	Synthetic Gag poly
32	52.4	5.1	810	24	ABK73201	Bacillus lichenifo
c 33	51.6	5.1	3412	20	AAX25772	S.erythraea erythr
34	51.6	5.1	3756	18	AAT72684	Sugar biosynthesis
35	51.6	5.1	27541	22	AAD17185	Streptomyces nours
36	51.6	5.1	125401	22	AAD17186	Streptomyces nours
37	51.2	5.0	1509	21	AAA51626	HIV codon-optimize
38	50.6	5.0	825	22	AAH00985	Mycobacterium aviu
39	50.6	5.0	23666	12	AAQ10190	Cephalosporin anti
40	50.4	4.9	1050	22	AAH41175	Mammalian Cre reco
41	50.4	4.9	1086	21	AAX89448	Branched chain ami
42	50.2	4.9	1835	21	AAA30798	DNA encoding maize
43	50.2	4.9	48328	22	AAF28540	Genomic fragment #
44	50	4.9	732	21	AAA39825	Synthetic DNA CAH-
45	50	4.9	29879	14	AAQ46806	eryA region of S.

ALIGNMENTS

RESULT 1
AAA94032
ID AAA94032 standard; DNA; 1020 BP.
XX
AC AAA94032;
XX
DT 15-JAN-2001 (first entry)
XX
DE Mycobacterium aurum stereoselective transamination catalyst gene.
XX
KW Transamination; optically active amino compound; isomer; ds.
XX
OS Mycobacterium aurum.
XX
FH Key Location/Qualifiers
FT CDS 1..1020
FT /*tag= a
FT /product= "stereoselective transamination catalyst"
XX
PN EP1038953-A1.
XX
PD 27-SEP-2000.
XX
PF 14-MAR-2000; 2000EP-0400701.
XX
PR 19-MAR-1999; 99JP-0075511.
PR 30-MAR-1999; 99JP-0088634.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Takashima Y, Mitsuda S, Wieser M;
XX

```
DR   WPI; 2000-629576/61.
DR   P-PSDB; AAB26100.
XX
PT   Novel Mycobacterium protein capable of catalyzing transamination stereo
PT   selectively, useful for producing optically active amino compounds,
PT   comprises a sequence of 339 amino acids   -
XX
PS   Claim 7; Page 33-38; 54pp; English.
XX
CC   The present sequence is the coding sequence for a protein from
CC   Mycobacterium aurum which acts as a biocatalyst. It is able to catalyse
CC   transamination stereoselectively to produce an optically active amino
CC   compound.
XX
SQ   Sequence 1020 BP; 198 A; 350 C; 313 G; 159 T; 0 other;

  Query Match             100.0%;  Score 1020;  DB 21;  Length 1020;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-186;
  Matches 1020;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60

Qy      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120

Qy     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180

Qy     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240

Qy     241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300

Qy     301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360

Qy     361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420

Qy     421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480

Qy     481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540

Qy     541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600

Qy     601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660

Qy     661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720

Qy     721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780

Qy     781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840

Qy     841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900

Qy     901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960

Qy     961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020


RESULT 2
AAA94040
ID   AAA94040 standard; DNA; 1020 BP.
XX
AC   AAA94040;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Mycobacterium aurum stereoselective transamination catalyst gene mutant.
XX
KW   Transamination; optically active amino compound; isomer; mutant;
KW   mutein; ds.
XX
OS   Mycobacterium aurum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1020
FT                   /*tag=  a
FT                   /product= "mutant stereoselective transamination
FT                   catalyst"
FT   misc_difference 4
FT                   /*tag=  b
FT                   /note= "wild-type A substituted by G"
XX
PN   EP1038953-A1.
XX
PD   27-SEP-2000.
XX
PF   14-MAR-2000; 2000EP-0400701.
XX
PR   19-MAR-1999;   99JP-0075511.
PR   30-MAR-1999;   99JP-0088634.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI   Takashima Y,  Mitsuda S,  Wieser M;
XX
DR   WPI; 2000-629576/61.
DR   P-PSDB; AAB26104.
XX
PT   Novel Mycobacterium protein capable of catalyzing transamination stereo
PT   selectively, useful for producing optically active amino compounds,
PT   comprises a sequence of 339 amino acids   -
XX
PS   Claim 8; Page -; 54pp; English.
XX
CC   The present sequence is a mutant version of the coding sequence for a
CC   protein from Mycobacterium aurum which acts as a biocatalyst. It is able
CC   to catalyse transamination stereoselectively to produce an optically
CC   active amino compound.
CC   Note: This sequence is not shown in the specification, but is derived
CC   from the Mycobacterium aurum wild-type sequence shown in SEQ ID NO: 2
CC   (see AAA94032)1.
XX
SQ   Sequence 1020 BP; 197 A; 350 C; 314 G; 159 T; 0 other;

  Query Match             99.8%;  Score 1018.4;  DB 21;  Length 1020;
```

```
  Best Local Similarity   99.9%;  Pred. No. 5e-186;
  Matches 1019;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60

Qy      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120

Qy     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180

Qy     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240

Qy     241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300

Qy     301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360

Qy     361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420

Qy     421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480

Qy     481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540

Qy     541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600

Qy     601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660

Qy     661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720

Qy     721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780

Qy     781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840

Qy     841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900

Qy     901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960

Qy     961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020

RESULT 3
```

```
AAV83675
ID   AAV83675 standard; DNA; 993 BP.
XX
AC   AAV83675;
XX
DT   26-FEB-1999  (first entry)
XX
DE   DNA encoding a protein with stereoselective transaminase activity.
XX
KW   Stereoselective; transaminase activity; optically active amine;
KW   R-chirality; ketone; synthetic intermediate; pharmaceutical;
KW   agrochemical; ds.
XX
OS   Arthrobacter sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..993
FT                   /*tag=  a
XX
PN   WO9848030-A1.
XX
PD   29-OCT-1998.
XX
PF   20-APR-1998;   98WO-JP01814.
XX
PR   23-APR-1997;   97JP-0121732.
XX
PA   (KANF ) KANEKA CORP.
XX
PI   Hasegawa J,  Ikenaka Y,  Iwasaki A,  Kizaki N,  Matsumoto K;
PI   Ogura M,  Yamada Y;
XX
DR   WPI; 1998-583664/49.
DR   P-PSDB; AAW87563.
XX
PT   Production of optically active amine(s) from ketone(s) and an amino
PT   donor - using a transaminase derived from an Arthrobacter species.
XX
PS   Claim 23; Page 57; 84pp; Japanese.
XX
CC   The present sequence encodes a protein with stereoselective
CC   transaminase activity. The transaminase is used to produce
CC   the optically active amines (having R-chirality) of the
CC   invention. A ketone is treated with a transaminase in the
CC   presence of a primary amine as amino donor to produce the
CC   amines. The chiral amines are synthetic intermediates for
CC   use as pharmaceuticals and agrochemicals.
XX
SQ   Sequence 993 BP; 185 A; 324 C; 297 G; 187 T; 0 other;
```

```
  Query Match           32.0%;  Score 326.6;  DB 19;  Length 993;
  Best Local Similarity   58.3%;  Pred. No. 9.7e-54;
  Matches  572;  Conservative    0;  Mismatches  409;  Indels    0;  Gaps    0;

Qy      40 GCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGC 99
           ||||      |||    | ||  |  | ||  | | | |      |||   || |||
Db      13 GCCGATACCTCCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGC 72

Qy     100 GACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAA 159
           |||||||||| ||  |  | | |||| |||| || || |  || ||||||||||| | |
Db      73 GACTACGAACTCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCA 132

Qy     160 TACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGAT 219
           | | ||||| |    || |||  |||||| ||||||||    || | |   || || ||
Db     133 TTCGTGCCGCCGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGAC 192

Qy     220 CTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTG 279
            | ||||||| |||   || || ||| |||| ||    ||||| |||| ||||||| |
Db     193 GTCACCTACACGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATC 252

Qy     280 GACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAA 339
           || ||  |  ||  |   |||     | ||||| |   | |   |  ||   ||||||||
Db     253 GAACGCCTCTTCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAA 312
```

```
Qy    340 CTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAAT 399
          ||   |||||  |      |   |||  ||| | |  |  |||| ||  | ||||||
Db    313 GTGAAGGAGATTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCC 372

Qy    400 CTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACC 459
          || |  | ||||| || ||| |    |      |||||||   ||| |  | |||| |  |
Db    373 GTGTCGATTACCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGC 432

Qy    460 CATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATC 519
          |  |||||||||||  || ||| |||| || | ||||  | | |||   | |        |
Db    433 CCGCAGGTGTACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGC 492

Qy    520 TTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGAC 579
            ||||     ||| || ||  | |      || |||||  ||   |||| | |  |||||
Db    493 GACGGCGTGCACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGAC 552

Qy    580 CCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGAC 639
          ||     |||||||| |||||||| ||||| | |   ||   |   ||  |  | |||
Db    553 CCTCAGGTCAAGAACTTCCAGTGGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGAC 612

Qy    640 CGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGC 699
          || ||        |  |  | |||||||||   ||         | ||||| ||  ||||
Db    613 CGCGGGTTCGAGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGG 672

Qy    700 TTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCG 759
          ||||||||   | |  ||||||||||||    ||  |  ||||   ||    ||| | ||
Db    673 TTCAACGTCGTCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCC 732

Qy    760 GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTG 819
          ||||| || || ||||| ||| |||  | |||||   | | ||   ||| || |  ||
Db    733 GGCATTACGCGGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTC 792

Qy    820 CGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCG 879
            |||| ||||   |  ||||||   ||||||||||||  || | |    ||| || |||
Db    793 GCCGACATCACGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCG 852

Qy    880 GGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGT 939
          ||||| ||    ||  ||  |    |||| ||| |  || |     |||| |  |||||
Db    853 GGCGGAGTGTGGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGC 912

Qy    940 CCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTG 999
          ||  | ||   | | |||      || | ||||| |||||  | ||||      | || ||
Db    913 CCCATCACCCAGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCCTCGTTG 972

Qy   1000 ATCGAAACGATCGAATACTGA 1020
          |     |  |  | ||||||
Db    973 CTTACGCCTGTGCAGTACTGA 993
```

RESULT 4
AAV83676
ID AAV83676 standard; DNA; 993 BP.
XX
AC AAV83676;
XX
DT 26-FEB-1999 (first entry)
XX
DE DNA encoding a protein with stereoselective transaminase activity.
XX
KW Stereoselective; transaminase activity; optically active amine;
KW R-chirality; ketone; synthetic intermediate; pharmaceutical;
KW agrochemical; ds.
XX
OS Arthrobacter sp.
XX
FH Key   Location/Qualifiers
FT CDS   1..978
FT   /*tag= a
XX
PN WO9848030-A1.
XX
PD 29-OCT-1998.
XX
PF 20-APR-1998; 98WO-JP01814.
XX
PR 23-APR-1997; 97JP-0121732.
XX
PA (KANF) KANEKA CORP.
XX
PI Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;
PI Ogura M, Yamada Y;
XX
DR WPI; 1998-583664/49.
DR P-PSDB; AAW87564.
XX
PT Production of optically active amine(s) from ketone(s) and an amino
PT donor - using a transaminase derived from an Arthrobacter species.
XX
PS Claim 23; Page 57; 84pp; Japanese.
XX
CC The present sequence encodes a protein with stereoselective
CC transaminase activity. The transaminase is used to produce
CC the optically active amines (having R-chirality) of the
CC invention. A ketone is treated with a transaminase in the
CC presence of a primary amine as amino donor to produce the
CC amines. The chiral amines are synthetic intermediates for
CC use as pharmaceuticals and agrochemicals.
XX
SQ Sequence 993 BP; 186 A; 324 C; 296 G; 187 T; 0 other;

Query Match 32.0%; Score 326.6; DB 19; Length 993;
Best Local Similarity 58.3%; Pred. No. 9.7e-54;
Matches 572; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

```
Qy     40 GCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGC 99
          ||||      |||    | ||  |  | ||  | | | |      |||   || |||
Db     13 GCCGATACCTCCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGC 72

Qy    100 GACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAA 159
          |||||||||||  ||  |  | | |||| |||| || ||  |  || ||||||||||| | |
Db     73 GACTACGAACTCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCA 132

Qy    160 TACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGAT 219
          | | |||||  |    || |||  |||||| ||||||||    || | |   || || ||
Db    133 TTCGTGCCGCCGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGAC 192

Qy    220 CTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTG 279
          | ||||||| |||   || || ||| |||| ||    ||||| |||| |||||||  |
Db    193 GTCACCTACACGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATC 252

Qy    280 GACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAA 339
          || ||  |  ||  |   |||     | ||||| |   | |   |  ||   ||||||||
Db    253 GAACGCCTCTTCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAA 312

Qy    340 CTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAAT 399
          ||   |||||  |      |   |||  ||| | |  |  |||| ||  | ||||||
Db    313 GTGAAGGAGATTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCC 372

Qy    400 CTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACC 459
          || |  | ||||| || ||| |    |      |||||||   ||| |  | |||| |  |
Db    373 GTGTCGATTACCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGC 432

Qy    460 CATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATC 519
          |  |||||||||||  || ||| |||| || | ||||  | | |||   | |        |
Db    433 CCGCAGGTGTACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGC 492

Qy    520 TTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGAC 579
            ||||     ||| || ||  | |      || |||||  ||   |||| | |  |||||
Db    493 GACGGCGTGCACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGAC 552

Qy    580 CCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGAC 639
          ||     |||||||| |||||||| ||||| | |   ||   |   ||  |  | |||
Db    553 CCTCAGGTCAAGAACTTCCAGTGGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGAC 612

Qy    640 CGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGC 699
```

```
Db    613 CGCGGGTTCGAGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGG 672

Qy    700 TTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCG 759

Db    673 TTCAACGTCGTCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCC 732

Qy    760 GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTG 819

Db    733 GGCATTACGCGGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTC 792

Qy    820 CGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCG 879

Db    793 GCCGACATCACGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCG 852

Qy    880 GGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGT 939

Db    853 GGCGGAGTGTGGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGC 912

Qy    940 CCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTG 999

Db    913 CCCATCACCCAGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCCTCGTTG 972

Qy   1000 ATCGAAACGATCGAATACTGA 1020

Db    973 CTTACGCCTGTGCAGTACTGA 993


RESULT 5
AAV73146
ID   AAV73146 standard; DNA; 978 BP.
XX
AC   AAV73146;
XX
DT   26-FEB-1999  (first entry)
XX
DE   DNA encoding a protein with stereoselective transaminase activity.
XX
KW   Stereoselective; transaminase activity; optically active amine;
KW   R-chirality; ketone; synthetic intermediate; pharmaceutical;
KW   agrochemical; ds.
XX
OS   Arthrobacter sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..978
FT                   /*tag=  a
XX
PN   WO9848030-A1.
XX
PD   29-OCT-1998.
XX
PF   20-APR-1998;   98WO-JP01814.
XX
PR   23-APR-1997;   97JP-0121732.
XX
PA   (KANF ) KANEKA CORP.
XX
PI   Hasegawa J,  Ikenaka Y,  Iwasaki A,  Kizaki N,  Matsumoto K;
PI   Ogura M,  Yamada Y;
XX
DR   WPI; 1998-583664/49.
DR   P-PSDB; AAW82241.
XX
PT   Production of optically active amine(s) from ketone(s) and an amino
PT   donor - using a transaminase derived from an Arthrobacter species.
XX
PS   Claim 23; Page 57; 84pp; Japanese.
XX
CC   The present sequence encodes a protein with stereoselective
CC   transaminase activity. The transaminase is used to produce
CC   the optically active amines (having R-chirality) of the
CC   invention. A ketone is treated with a transaminase in the
CC   presence of a primary amine as amino donor to produce the
CC   amines. The chiral amines are synthetic intermediates for
CC   use as pharmaceuticals and agrochemicals.
XX
SQ   Sequence 978 BP; 183 A; 319 C; 292 G; 184 T; 0 other;

  Query Match             32.0%;  Score 326.2;  DB 19;  Length 978;
  Best Local Similarity   58.5%;  Pred. No. 1.2e-53;
  Matches  568;  Conservative    0;  Mismatches  403;  Indels    0;  Gaps    0;

Qy     50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109

Db      8 CCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGCGACTACGAAC 67

Qy    110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169

Db     68 TCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCATTCGTGCCGC 127

Qy    170 CCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACA 229

Db    128 CGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGACGTCACCTACA 187

Qy    230 CCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGC 289

Db    188 CGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATCGAACGCCTCT 247

Qy    290 TCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGA 349

Db    248 TCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAAGTGAAGGAGA 307

Qy    350 TCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCA 409

Db    308 TTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCCGTGTCGATTA 367

Qy    410 CCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGT 469

Db    368 CCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGCCCGCAGGTGT 427

Qy    470 ACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCA 529

Db    428 ACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGCGACGGCGTGC 487

Qy    530 CCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCA 589

Db    488 ACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGACCCTCAGGTCA 547

Qy    590 AGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGC 649

Db    548 AGAACTTCCAGTGGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGACCGCGGGTTCG 607

Qy    650 GCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGT 709

Db    608 AGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGGTTCAACGTCG 667

Qy    710 GCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCC 769

Db    668 TCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCCGGCATTACGC 727

Qy    770 GTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCA 829

Db    728 GGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTCGCCGACATCA 787

Qy    830 CCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCA 889

Db    788 CGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCGGGCGGAGTGT 847

Qy    890 CACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGG 949

Db    848 GGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGCCCCATCACCC 907

Qy    950 TGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGA 1009

Db    908 AGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCCTCGTTGCTTACGCCTG 967
```

```
Qy   1010 TCGAATACTGA 1020
          |  | ||||||
Db    968 TGCAGTACTGA 978


RESULT 6
AAV73156
ID   AAV73156 standard; DNA; 1574 BP.
XX
AC   AAV73156;
XX
DT   26-FEB-1999  (first entry)
XX
DE   PstI fragment of a stereoselective transaminase encoding genomic DNA.
XX
KW   Stereoselective; transaminase activity; optically active amine;
KW   R-chirality; ketone; synthetic intermediate; pharmaceutical;
KW   agrochemical; ss.
XX
OS   Arthrobacter sp.
XX
PN   WO9848030-A1.
XX
PD   29-OCT-1998.
XX
PF   20-APR-1998;   98WO-JP01814.
XX
PR   23-APR-1997;   97JP-0121732.
XX
PA   (KANF ) KANEKA CORP.
XX
PI   Hasegawa J,  Ikenaka Y,  Iwasaki A,  Kizaki N,  Matsumoto K;
PI   Ogura M,  Yamada Y;
XX
DR   WPI; 1998-583664/49.
XX
PT   Production of optically active amine(s) from ketone(s) and an amino
PT   donor - using a transaminase derived from an Arthrobacter species.
XX
PS   Example 21; Page 63; 84pp; Japanese.
XX
CC   The present sequence represents a 1.6 kb PstI fragment of
CC   genomic DNA encoding a protein with stereoselective
CC   transaminase activity. The transaminase is used to produce
CC   the optically active amines (having R-chirality) of the
CC   invention. A ketone is treated with a transaminase in the
CC   presence of a primary amine as amino donor to produce the
CC   amines. The chiral amines are synthetic intermediates for
CC   use as pharmaceuticals and agrochemicals.
XX
SQ   Sequence 1574 BP; 285 A; 530 C; 476 G; 283 T; 0 other;

  Query Match             31.9%;  Score 325;  DB 19;  Length 1574;
  Best Local Similarity   58.2%;  Pred. No. 2e-53;
  Matches  571;  Conservative    0;  Mismatches  410;  Indels    0;  Gaps    0;

Qy     40 GCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGC 99
          ||||      |||    | ||  |  | ||  | | | |  |      |||   || |||
Db    398 GCCGATACCTCCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGC 457

Qy    100 GACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAA 159
          |||||||||||  ||  |  | | |||| |||| || || |  || ||||||||||| | |
Db    458 GACTACGAACTCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCA 517

Qy    160 TACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGAT 219
          | | ||||| |    || |||  |||||| ||||||||    || | |   || || ||
Db    518 TTCGTGCCGCCGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGAC 577

Qy    220 CTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTG 279
          | ||||||||| |||   || || ||| |||| ||    ||||| |||| |||||||  |
Db    578 GTCACCTACACGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATC 637

Qy    280 GACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAA 339
          || ||  |  ||  |   |||      | ||||| |    | |    |  ||   ||||||||
Db    638 GAACGCCTCTTCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAA 697

Qy    340 CTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAAT 399
          ||   |||||  |      |  |||  ||| | |  |  |||| ||  | ||||||
Db    698 GTGAAGGAGATTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCC 757

Qy    400 CTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACC 459
          || |  | ||||| || ||| |   |      |||||||  ||| |  | |||| |  |
Db    758 GTGTCGATTACCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGC 817

Qy    460 CATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATC 519
          |  ||||||||||| || ||| |||| || | ||||  | | |||   | |        |
Db    818 CCGCAGGTGTACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGC 877

Qy    520 TTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGAC 579
            ||||     ||| || ||  | |      || |||||  ||   |||| | |  |||||
Db    878 GACGGCGTGCACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGAC 937

Qy    580 CCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGAC 639
          ||     ||||||||| ||||||||| ||||| | |   ||   |   ||  |  | |||
Db    938 CCTCAGGTCAAGAACTTCCAGTGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGAC 997

Qy    640 CGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGC 699
          || ||         |  |  | |||||||||   ||         | ||||| ||  ||||
Db    998 CGCGGGTTCGAGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGG 1057

Qy    700 TTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCG 759
          ||||||||   | |  ||||||||||||     ||  |  ||||   ||    ||| | ||
Db   1058 TTCAACGTCGTCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCC 1117

Qy    760 GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTG 819
          |||||  ||  ||  |||||  |||  ||||   | |||||    | | ||   ||| || |  ||
Db   1118 GGCATTACGCGGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTC 1177

Qy    820 CGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCG 879
            ||||  ||||   |  ||||||   ||||||||||||  ||  | |    ||| || |||
Db   1178 GCCGACATCACGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCG 1237

Qy    880 GGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGT 939
          |||||  ||    ||  ||  |    |||| ||| |  || |      |||| |  |||||
Db   1238 GGCGGAGTGTGGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGC 1297

Qy    940 CCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTG 999
          ||  |  ||    | | |||       || | ||||| |||||  | ||||        || ||
Db   1298 CCCATCACCCAGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCGTCGTTG 1357

Qy   1000 ATCGAAACGATCGAATACTGA 1020
          |     |  |  | ||||||
Db   1358 CTTACGCCTGTGCAGTACTGA 1378


RESULT 7
AAT68715
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag=  a
FT   CDS             14056..14136
```

```
FT                   /*tag=  b
FT   CDS             14148..15827
FT                   /*tag=  c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US20119.
XX
PR   19-DEC-1995;   95US-0008847.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
DR   P-PSDB; AAW19629-30 AND AAW00918.
XX
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also AAW19629-30 and AAW00918).  It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain.  Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain.  The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library.  A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.
XX
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

  Query Match             6.5%;  Score 66.2;  DB 18;  Length 15872;
  Best Local Similarity   44.4%;  Pred. No. 0.00063;
  Matches  358;  Conservative    0;  Mismatches  443;  Indels    6;  Gaps    2;

Qy      50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109
           || ||||| ||  ||| | |  ||  | |  |   || ||||    ||||| |  | |
Db     765 CCCGCGCCCTCGCCGACGGCGACCGTGTCCACGGCGTCATCCGCGCCAGCGCCGTCAACA 824

Qy     110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169
             |||    ||| ||||     |  ||    |  | ||  ||     ||  |  ||| |
Db     825 ACGACGGAGCCACCCCGGGTCTCACCGTGCCCAGCAGGGCCGCCCAGGAGAAGGTGCTGC 884

Qy     170 CCGAAG---AAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCT 226
           |||  |   |   ||||    || |   |    |||   ||    |   ||| ||   |
Db     885 GCGAGGCGTACCGGAAGGCGGCCCTGGACCCGTCCGCCGTCCAGTACGTCGAACTCCACG 944

Qy     227 ACACCGTCGCGCATGTATGGCACGGCA---ACATCTTCCGGCTCGGCGACCACCTGGACC 283
           |||||   | |  ||  |  ||  ||   |   |  |  ||||||| |  ||| | |
Db     945 GCACCGGAACCCCCGTCGGCGACCCCATCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCT 1004

Qy     284 GGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGG 343
           |  || |  || ||         |||| | | | | | |    | |  || | ||  |
Db    1005 CGGCGCGCCCCGCGGACGAACCCCTGCTCGTCGGCTCGGCCAAGACGAACGTCGGGCACC 1064

Qy     344 CCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGA 403
           |||   | ||        |||     |     ||  || |    || |||     |
Db    1065 TCGAAGGCGCCGCCGGCATCGTCGGCCTCATCAAGACGCTCCTCGCGCTCGGCCCGCGCC 1124

Qy     404 CCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATC 463
             || |   | |   |      || | |  | |   |||| |  |     |||||
Db    1125 GGATCCCGGCGAGCCTCAACTTCCGTACGCCCCACCCGGACATCCCGCTCGACACCCTCG 1184

Qy     464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523
```

```
           | |  || |   || |  || |     |||| |   ||||  |  ||| |   ||  ||
Db    1185 GGCTCGACGTGCCCGACGGCCTGCGGGAGTGGCCGCACCCGGACCGCGAACTCCTCGCCG 1244

Qy     524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583
           ||  || | || |||    | ||     |  |  |  || |  |  || ||  | |||
Db    1245 GCGTCAGCTCGTTCGGCATGGGCGGCACCAACGCCCACGTCGTCCTCAGCGAAGGCCCCG 1304

Qy     584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643
           ||        | | |||   ||          |   ||    ||||   |  || |
Db    1305 CCCAGGGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCCGTCGACAGCGGGGCCGCAC 1364

Qy     644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703
             |   |  ||  | |   | || |   ||   |      ||| | |  | | || |
Db    1365 TGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGCGCCCAGGCCCGGCGCCTGC 1424

Qy     704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763
           ||| |  | |||    ||||    |||| ||  || |  | |     |  || || || ||
Db    1425 ACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCACTCGCCCGGTCGCTGGTCA 1484

Qy     764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823
           ||||||||  |    |     | | | | | |    | | |   ||   | |||  | |
Db    1485 CCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCC 1544

Qy     824 ACGTCACCAGCCGTGAACTCTACGACG 850
           || |  |  | | |  ||| | || ||
Db    1545 TCGACGGCCTCGGCGCCCTCGCCGCCG 1571


RESULT 8
AAZ87283
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag=  a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056..14151
FT                   /*tag=  b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167..15827
FT                   /*tag=  c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
```

PT synthesis of methymycin and pikromycin -
XX
PS Example 3; Figure 23; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants. The
CC present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC sequence, designated vep ORF 1 in the specification, which actually
CC contains 3 open reading frames, which encode proteins AAY77177-Y77178 and
CC AAY77199. The vep ORF 1 protein is defined in the specification as a PHA
CC monomer synthase. .
XX
SQ Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

Query Match 6.5%; Score 66.2; DB 21; Length 15872;
Best Local Similarity 44.4%; Pred. No. 0.00063;
Matches 358; Conservative 0; Mismatches 443; Indels 6; Gaps 2;

```
Qy      50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109
           || ||||| ||  ||| | |  ||  | |   |   || ||||    ||||| |  | |
Db     765 CCCGCGCCCTCGCCGACGGCGACCGTGTCCACGGCGTCATCCGCGCCAGCGCCGTCAACA 824

Qy     110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169
             |||    ||| ||||     |  ||    |  | ||  ||     ||  |  ||| |
Db     825 ACGACGGAGCCACCCCGGGTCTCACCGTGCCCAGCAGGGCCGCCCAGGAGAAGGTGCTGC 884

Qy     170 CCGAAG---AAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCT 226
            ||| |   |   ||||    || |   |    |||   ||    |   ||| ||   |
Db     885 GCGAGGCGTACCGGAAGGCGGCCCTGGACCCGTCCGCCGTCCAGTACGTCGAACTCCACG 944

Qy     227 ACACCGTCGCGCATGTATGGCACGGCA---ACATCTTCCGGCTCGGCGACCACCTGGACC 283
           |||||   | |  ||  |  ||  ||   |   | |  ||||||||  |  ||| | |
Db     945 GCACCGGAACCCCCGTCGGCGACCCCATCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCT 1004

Qy     284 GGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGG 343
           |  || |  || ||        |||| | | | | | |   | |  || | ||  |
Db    1005 CGGCGCGCCCCGCGGACGAACCCCTGCTCGTCGGCTCGGCCAAGACGAACGTCGGGCACC 1064

Qy     344 CCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGA 403
           |||   | ||       |||     |     ||  || |    || |||     |
Db    1065 TCGAAGGCGCCGCCGGCATCGTCGGCCTCATCAAGACGCTCCTCGCGCTCGGCCGGCGCC 1124

Qy     404 CCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATC 463
             || |   | |   |      || | |  | |   |||| |  |       |||||
Db    1125 GGATCCCGGCGAGCCTCAACTTCCGTACGCCCCACCCGGACATCCCGCTCGACACCCTCG 1184

Qy     464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523
           | |  || |   || |  || |     |||| |   ||||  |  ||| |   ||  ||
Db    1185 GGCTCGACGTGCCCGACGGCCTGCGGGAGTGGCCGCACCCGGACCGCGAACTCCTCGCCG 1244

Qy     524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583
           ||  || | || |||    | ||      |  |  |  || |  |  || ||  | |||
Db    1245 GCGTCAGCTCGTTCGGCATGGGCGGCACCAACGCCCACGTCGTCCTCAGCGAAGGCCCCG 1304

Qy     584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643
           ||        | | |||   ||          |   ||    ||||   |  || |
Db    1305 CCCAGGGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCCGTCGACAGCGGGGCCGCAC 1364

Qy     644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703
           |   |  ||  | |   | || |   ||   |       ||| | |  | | || |
Db    1365 TGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGCGCCCAGGCCCGGCGCCTGC 1424

Qy     704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763
           ||| |  | |||    ||||    |||| ||  || |  | |    |  || || || ||
Db    1425 ACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCACTCGCCCGGTCGCTGGTCA 1484

Qy     764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823
           |||||||| |    |    | | | | | |    | | |   ||   | |||  | |
Db    1485 CCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCC 1544

Qy     824 ACGTCACCAGCCGTGAACTCTACGACG 850
           || |  |  | | |  |||  || ||
Db    1545 TCGACGGCCTCGGCGCCCTCGCCGCCG 1571
```

RESULT 9
AAA58471
ID AAA58471 standard; DNA; 58857 BP.
XX
AC AAA58471;
XX
DT 31-OCT-2000 (first entry)
XX
DE Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS Streptomyces verticillus.
XX

FH	Key	Location/Qualifiers
FT	CDS	223..564
FT		/*tag= a
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 30; encodes AAB07556"
FT	CDS	561..2309
FT		/*tag= b
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 29; encodes AAB07557"
FT	CDS	2767..3486
FT		/*tag= c
FT		/note= "ORF 28; encodes AAB07558"
FT	CDS	3527..5593
FT		/*tag= d
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 27; encodes AAB07559"
FT	CDS	5806..12294
FT		/*tag= e
FT		/note= "ORF 26; encodes AAB07560"
FT	CDS	12291..15491
FT		/*tag= f
FT		/note= "ORF 25; encodes AAB07561"
FT	CDS	15488..21013
FT		/*tag= g
FT		/note= "ORF 24; encodes AAB07562"
FT	CDS	21010..24666
FT		/*tag= h
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 23; encodes AAB07563"
FT	CDS	24663..32690
FT		/*tag= i
FT		/note= "ORF 22; encodes AAB07564"

```
FT   CDS             32893..34830
FT                   /*tag=  j
FT                   /note= "ORF 21; encodes AAB07565"
FT   CDS             34827..35804
FT                   /*tag=  k
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 20; encodes AAB07566"
FT   CDS             35818..37302
FT                   /*tag=  l
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 19; encodes AAB07567"
FT   CDS             37299..39215
FT                   /*tag=  m
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 18; encodes AAB07568"
FT   CDS             39301..47181
FT                   /*tag=  n
FT                   /note= "ORF 17; encodes AAB07569"
FT   CDS             47178..49985
FT                   /*tag=  o
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 16; encodes AAB07570"
FT   CDS             49982..51001
FT                   /*tag=  p
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 15; encodes AAB07571"
FT   CDS             50998..52386
FT                   /*tag=  q
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 14; encodes AAB07572"
FT   CDS             52383..52946
FT                   /*tag=  r
FT                   /note= "ORF 13; encodes AAB07573"
FT   CDS             53018..54190
FT                   /*tag=  s
FT                   /note= "ORF 12; encodes AAB07574"
FT   CDS             54187..55824
FT                   /*tag=  t
FT                   /note= "ORF 11; encodes AAB07575"
FT   CDS             55821..56093
FT                   /*tag=  u
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 10; encodes AAB07576"
FT   CDS             56090..57586
FT                   /*tag=  v
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 9; encodes AAB07577"
FT   CDS             57583..58857
FT                   /*tag=  w
FT                   /note= "ORF 8; encodes AAB07578"
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00445.
XX
PR   06-JAN-1999;   99US-0115435.
PR   05-FEB-1999;   99US-0118848.
PR   05-JAN-2000; 2000US-0477962.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Shen B,  Du L,  Sanchez C,  Chen M,  Edwards DJ;
XX
DR   WPI; 2000-465974/40.
DR   P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR   AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR   AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR   AAB07576, AAB07577, AAB07578.
XX
PT   New bleomycin gene cluster components useful for peptide and/or
PT   polyketide metabolites, especially bleomycin, production and for
PT   chemically modifying biological molecules -
XX
PS   Claim 8; Page 97-136; 162pp; English.
XX
CC   The present sequence represents the BLM (Bleomycin) gene cluster,
CC   containing open reading frames (ORFs) 8-30. The proteins encoded
CC   by the gene cluster are useful for producing peptides and/or polyketide
CC   metabolites, especially bleomycin or bleomycin analogues. They are
CC   also useful for chemically modifying biological molecules to produce
CC   branched methyl groups, and for coupling amino acids and fatty
CC   acids. They may be reacted with an apo-carrier protein and coenzyme A
CC   to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC   domains can be used individually or collectively to produce
CC   thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC   microbial metabolites. The BLM gene cluster may also be used to produce
CC   sugars.
XX
SQ   Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;

  Query Match             6.3%;  Score 64.2;  DB 21;  Length 58857;
  Best Local Similarity   42.4%;  Pred. No. 0.0016;
  Matches  345;  Conservative    0;  Mismatches  468;  Indels    0;  Gaps    0;

Qy      40 GCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGC 99
           | | ||  |  |  || |   | ||  |   | || ||| || |     |||| ||| ||
Db   12918 GTCTTCCGGGGCACCGGCCGGCCCGCCGGACCGCCCGCCCGCGCCACCGTCCACTACGGC 12977

Qy     100 GACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAA 159
           | |||| | || | |  ||       |||  | |||    ||||   | ||    ||
Db   12978 ACCGACGATCCGGCCCCGACCGCGGACTTCTGCCGCGAGCACGCCGTCACCGGCTACGTG 13037

Qy     160 TACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGAT 219
             ||||  || |   ||     |   |   |    |||||||        |    |
Db   13038 CTGCTGCTCGCGGCCCTCGCCTGCCTGGTCGCCCGGTACACCGGCCGGACGGACGTGGTG 13097

Qy     220 CTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTG 279
           |   ||   ||||||  |       | ||  | |     ||     || || ||   ||
Db   13098 ATCGGCTCACCCGTCGGACTGCGCGAGGACCCCGAAGGGCTCGCCACCGTCGGCCCGATG 13157

Qy     280 GACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAA 339
           |    ||||       |      | ||||| |   ||| |||| | | || ||| |
Db   13158 CTCAACCTGCTGCCGCTGCGCCTCCGGCTGCACGGCGACCCCGGCTTCGGCGAGGTCCTG 13217

Qy     340 CTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAAT 399
           |          | |  |  ||   || ||  ||| |     |  | ||||| | |
Db   13218 GCCCGCACCCGGGAGACGCTGCTCGGCGCGCTGGAGCACCGCACCACACCGTTCGAGGAC 13277

Qy     400 CTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACC 459
           |   || | |   |||   ||     |  | || ||  |    ||| | | || ||  |
Db   13278 ATCGTCGACGCGGTGGGCGCCGACCGGGACCCGGACGTCAGCCCCCTCTTCCAGATCCTC 13337

Qy     460 CATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATC 519
                 |   | | |   |||   |  ||   |   |   |  |||       |
Db   13338 TTCGCCCACGAACGCCCCCCGGCCCCACCCGCGTTACCGGGCGTCCGTGCCCGCGTCGTA 13397

Qy     520 TTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCCGGCCGCAACACCGTCGAC 579
           || | || |  ||  ||     | | |  ||  |  | ||| |  | | ||   ||||
Db   13398 CCCGTCCCCGCTCCGGCCGCCAAGTACGAGCTCGCCGTCACCGCCACCGAGACGCCCGAC 13457

Qy     580 CCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGAC 639
           |  |      || |  | ||  ||   |       ||  |       ||||  |  | |
Db   13458 GGGCTCCGGCTGATCGTCGAGGCGGAGCACGGACACGGGGAACCGGCCGAACTCGCCGCC 13517

Qy     640 CGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGC 699
             |     | |  ||   |||||||  | | | |     | ||| | | ||    |||
Db   13518 TTCGCCCGCCACTTCGGCGTCCTGCTGGCCGCCGGGGTCCGCGCGCCGGACACACCGCTG 13577

Qy     700 TTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCG 759
           |  | ||     |  |||  ||||   ||| |  |   |   ||   | | ||   ||
Db   13578 AGCCGCCTGCCGCTGCTCACCGACGAGGAGCGGCGCCGGCTCACCGACACCACGGCCCCC 13637

Qy     760 GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTG 819
             ||| | | |   || |    | |   |      ||| | ||| |   ||  || ||
```

```
Db  13638 CGCACCGCGCCGGAGGCCCCCTACCGCCCCCTGCACCGGCTGGTCGAGGAGTCCGCCGCC 13697

Qy    820 CGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
          |||      |  |  ||  |   ||  || || |
Db  13698 CGCCGGCCCGACGCCCTGGCGGTCGTCGGCGGC 13730


RESULT 10
AAQ64204
ID   AAQ64204 standard; cDNA; 1208 BP.
XX
AC   AAQ64204;
XX
DT   18-NOV-1994  (first entry)
XX
DE   snaC gene encoding enzyme in streptogramin biosynthetic pathway.
XX
KW   Antibiotic; streptogramin; snaA; snaB; snaC; biosynthesis; enzyme;
KW   biosynthetic pathway; Streptomyces pristinaespiralis; ds.
XX
OS   Streptomyces pristinaespiralis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1209
FT                   /*tag=  a
XX
PN   FR2696189-A.
XX
PD   01-APR-1994.
XX
PF   25-SEP-1992;   92FR-0011441.
XX
PR   25-SEP-1992;   92FR-0011441.
XX
PA   (RHON ) RHONE POULENC RORER SA.
XX
PI   Blanc V,  Blanche F,  Crouzet J,  Jacques N,  Lacroix P;
PI   Thibaut D,  Zagorec M;
XX
DR   WPI; 1994-128286/16.
DR   P-PSDB; AAR54204.
XX
PT   DNA involved in streptogramin antibiotic biosynthesis - for
PT   prodn. or bio-conversion of streptogramin(s) or prodn. of
PT   streptogramin intermediates, derivs. or hybrid antibiotics
XX
PS   Claim 2; Page 54-55; 83pp; French.
XX
CC   The snaC gene product is involved in the biosynthesis of
CC   streptogramins, antibiotics active against Gram-positive bacteria.
CC   The identification of the sequences encoding the enzymes involved
CC   in the biosynthetic pathway means that they can be isolated and
CC   manipulated.  Mutant microorganisms in which a step in the
CC   streptogramin biosynthetic pathway is blocked can be cultured to
CC   produce streptogramin intermediates, which may later be converted
CC   to streptogramin derivatives.  Recombinant cells may also be used
CC   for the bioconversion of streptogramins from one form to another or
CC   for the production of hybrid antibiotics.
XX
SQ   Sequence 1208 BP; 190 A; 500 C; 360 G; 158 T; 0 other;

  Query Match             5.8%;  Score 59;  DB 15;  Length 1208;
  Best Local Similarity   42.3%;  Pred. No. 0.014;
  Matches  390;  Conservative    0;  Mismatches  530;  Indels    3;  Gaps    1;

Qy     16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75
          |||  || |  |  |  |||| ||  ||  || |||| | |   |  || || |    |
Db     73 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 132

Qy     76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135
            | |      ||||  |  |  |  ||    || |||| |||   ||| |||   |
Db    133 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 192

Qy    136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195
          || |     | ||   | |     |||| |   ||   | |     |  | |||    |
Db    193 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 252

Qy    196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255
          |  |||   |          ||        || || || ||||||      | ||  |  |
Db    253 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 312

Qy    256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
             |    | ||| | ||  |  |||  |    | || ||    | ||||   || |
Db    313 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 372

Qy    316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372
          | | ||    |  | ||| |    ||   | | |   |||| |  ||  |   ||| ||
Db    373 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 432

Qy    373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
          || ||||| |      | |||   | |     |    ||||     ||     ||||||
Db    433 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 492

Qy    433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
          | ||  |  | ||  | |  || | ||  |     |  |     | ||||| ||||| |
Db    493 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 552

Qy    493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
          |   |  ||||     |  | | | | ||| |  | ||  |         |  |||  ||
Db    553 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCCACATC 612

Qy    553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
          || || |  ||    ||||  ||||      |    |  | | ||  |       |
Db    613 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 672

Qy    613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
          |   |||     ||    ||||  |      |    | |   | ||| || || ||
Db    673 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 732

Qy    673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
          |  | |  | |     ||       |   ||    |  | | |  | | |||||  |
Db    733 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 792

Qy    733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
          |  |    |      ||   || |  ||  |  | ||    || | | ||    || |
Db    793 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 852

Qy    793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
              | || |  |    || || |  ||  |  | |    |     ||| || |
Db    853 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 912

Qy    853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
          | || |    | ||   |  |  || || |  |||| |      || |     ||  |
Db    913 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 972

Qy    913 GAGGCCGTGGGCAACGGCGAGCC 935
          || ||       || |||| |
Db    973 CCGGTCGGCCTGTTCGTCGAGAC 995


RESULT 11
ABK91617
ID   ABK91617 standard; DNA; 9169 BP.
XX
AC   ABK91617;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Modified HIV protein-encoding plasmid DNA #169.
XX
KW   HIV; human immunodeficiency virus; gene; ds; circular; cyclic; anti-HIV;
KW   Env; Gag; Pol; Nef; genetic immunisation; protein immunisation; AIDS;
KW   acquired immunodeficiency syndrome; cytotoxic T lymphocyte; CTL.
XX
OS   Synthetic.
XX
```

```
PN   WO200232943-A2.
XX
PD   25-APR-2002.
XX
PF   14-AUG-2001; 2001WO-US25721.
XX
PR   14-AUG-2000; 2000US-225097P.
PR   14-NOV-2000; 2000US-252115P.
PR   28-MAR-2001; 2001US-279257P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (CHAD/) CHADRABARTI B K.
XX
PI   Nabel GJ,  Huang Y;
XX
DR   WPI; 2002-452382/48.
XX
PT   New modified human immunodeficiency virus Env, Gag, Pol of Nef DNA and
PT   its encoded protein, useful as vaccines for genetic or protein
PT   immunisation for acquired immunodeficiency syndrome or human
PT   immunodeficiency virus infection -
XX
PS   Disclosure; Page 769-772; 794pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a modified HIV
CC   (human immunodeficiency virus) protein. The nucleic acid molecule and its
CC   encoded protein (modified HIV Env, Gag, Pol or Nef) are useful for
CC   genetic or protein immunisation to a host, respectively. In particular
CC   these are useful for ameliorating the symptoms of acquired
CC   immunodeficiency syndrome (AIDS) or HIV infection and generating an
CC   antibody or cytotoxic T lymphocyte (CTL) response against native HIV Env,
CC   Gag, Pol or Nef. Furthermore, the nucleic acids, proteins and subsequent
CC   compositions are useful for treating or preventing HIV infections or
CC   AIDS. Sequences ABK91449-ABK91624 represent modified HIV protein-encoding
CC   plasmid DNA sequences of the invention.
XX
SQ   Sequence 9169 BP; 2227 A; 2668 C; 2505 G; 1769 T; 0 other;

  Query Match             5.7%;  Score 58.4;  DB 24;  Length 9169;
  Best Local Similarity   49.4%;  Pred. No. 0.019;
  Matches  152;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy    315 GGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTC 374
          |||| ||  | |   |  |||| |  | |  ||||||| || ||| ||  |  || | |
Db   2164 GGACACCAAGGAGGCCCTGGACAAGATCGAGGAGATCAAGAACAAGAGCAAGCAGAAGAC 2223

Qy    375 GCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGG 434
          |||| |  ||   |   |          | |  || |  | |  |      | | ||  |
Db   2224 CCAGCAGGCCGCCGCCGACACCGGCAACAGCAGCAACGTGAGCCAGAACTACCCCATCGT 2283

Qy    435 CGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTG 494
          |||| | || |  | || | | ||| |||      | |  |  || || | ||   ||
Db   2284 GCAGAACGCCCAGGGCCAGATGATCCACCAGAGCCTGAGCCCCCGCACCCTGAACGCCTG 2343

Qy    495 GGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCG 554
          ||       |  ||  ||| ||  |||| || ||   | ||||     |  |   | ||
Db   2344 GGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTCAGCGCCCT 2403

Qy    555 CCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCAC 614
           |||  |||  || | ||   |||| || ||| | |    ||     ||| |  | |
Db   2404 GAGCGAGGGCGCCACCCCCCAGGACCTGAACATGATGCTGAACATCGTGGGCGGCCACCA 2463

Qy    615 CGCAGCCA 622
          || ||||
Db   2464 GGCCGCCA 2471


RESULT 12
ABK91624
ID   ABK91624 standard; DNA; 9785 BP.
XX
AC   ABK91624;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Modified HIV protein-encoding plasmid DNA #176.
XX
KW   HIV; human immunodeficiency virus; gene; ds; circular; cyclic; anti-HIV;
KW   Env; Gag; Pol; Nef; genetic immunisation; protein immunisation; AIDS;
KW   acquired immunodeficiency syndrome; cytotoxic T lymphocyte; CTL.
XX
OS   Synthetic.
XX
PN   WO200232943-A2.
XX
PD   25-APR-2002.
XX
PF   14-AUG-2001; 2001WO-US25721.
XX
PR   14-AUG-2000; 2000US-225097P.
PR   14-NOV-2000; 2000US-252115P.
PR   28-MAR-2001; 2001US-279257P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (CHAD/) CHADRABARTI B K.
XX
PI   Nabel GJ,  Huang Y;
XX
DR   WPI; 2002-452382/48.
XX
PT   New modified human immunodeficiency virus Env, Gag, Pol of Nef DNA and
PT   its encoded protein, useful as vaccines for genetic or protein
PT   immunisation for acquired immunodeficiency syndrome or human
PT   immunodeficiency virus infection -
XX
PS   Disclosure; Page 791-794; 794pp; English.
XX
CC   The invention relates to a nucleic acid molecule encoding a modified HIV
CC   (human immunodeficiency virus) protein. The nucleic acid molecule and its
CC   encoded protein (modified HIV Env, Gag, Pol or Nef) are useful for
CC   genetic or protein immunisation to a host, respectively. In particular
CC   these are useful for ameliorating the symptoms of acquired
CC   immunodeficiency syndrome (AIDS) or HIV infection and generating an
CC   antibody or cytotoxic T lymphocyte (CTL) response against native HIV Env,
CC   Gag, Pol or Nef. Furthermore, the nucleic acids, proteins and subsequent
CC   compositions are useful for treating or preventing HIV infections or
CC   AIDS. Sequences ABK91449-ABK91624 represent modified HIV protein-encoding
CC   plasmid DNA sequences of the invention.
XX
SQ   Sequence 9785 BP; 2360 A; 2843 C; 2731 G; 1851 T; 0 other;

  Query Match             5.7%;  Score 58.4;  DB 24;  Length 9785;
  Best Local Similarity   49.4%;  Pred. No. 0.019;
  Matches  152;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy    315 GGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTC 374
          |||| ||  | |   |  |||| |  | |  ||||||| || ||| ||  |  || | |
Db   2164 GGACACCAAGGAGGCCCTGGACAAGATCGAGGAGATCAAGAACAAGAGCAAGCAGAAGAC 2223

Qy    375 GCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGG 434
          |||| |  ||   |   |          | |  || |  | |  |      | | ||  |
Db   2224 CCAGCAGGCCGCCGCCGACACCGGCAACAGCAGCAACGTGAGCCAGAACTACCCCATCGT 2283

Qy    435 CGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTG 494
          |||| | || |  | || | | ||| |||      | |  |  || || | ||   ||
Db   2284 GCAGAACGCCCAGGGCCAGATGATCCACCAGAGCCTGAGCCCCCGCACCCTGAACGCCTG 2343

Qy    495 GGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCG 554
          ||       |  ||  ||| ||  |||| || ||   | ||||     |  |   | ||
Db   2344 GGTGAAGGTGATCGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTCAGCGCCCT 2403

Qy    555 CCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCAC 614
           |||  |||  || | ||   |||| || ||| | |    ||     ||| |  | |
Db   2404 GAGCGAGGGCGCCACCCCCCAGGACCTGAACATGATGCTGAACATCGTGGGCGGCCACCA 2463

Qy    615 CGCAGCCA 622
```

```
                   || ||||
Db    2464 GGCCGCCA 2471


RESULT 13
AAQ63293
ID   AAQ63293 standard; DNA; 12588 BP.
XX
AC   AAQ63293;
XX
DT   04-JAN-1995  (first entry)
XX
DE   Sequence encoding mannuronan C-5-epimerase genes.
XX
KW   Mannuronan; epimerase; alginates; immunostimulant;
KW   cell immobilisation; Azotobacter vinelandii; guluronic acid;
KW   mannuronic acid; ss.
XX
OS   Azotobacter vinelandii strain E.
XX
FH   Key             Location/Qualifiers
FT   CDS             290..1951
FT                   /*tag= a
FT                   /product= Epimerase 4.
FT   CDS             2227..6438
FT                   /*tag= b
FT                   /product= Epimerase 1.
FT   CDS             6702..9695
FT                   /*tag= c
FT                   /product= Epimerase 2.
FT   CDS             9973..12588
FT                   /*tag= d
FT                   /product= First part of epimerase 3.
FT                   /note= "The specification states the sequence
FT                          length to be 12411 base pairs but the actual
FT                          sequence reproduced in the specification is
FT                          12588 base pairs.  The specification also
FT                          states that the first part of epimerase 3
FT                          terminates at position 12411 although in the
FT                          sequence there is no stop codon found at this
FT                          position and the given nucleotide sequence
FT                          is decoded in its entirity to give the
FT                          corresponding amino acid sequence."
XX
PN   WO9409124-A.
XX
PD   28-APR-1994.
XX
PF   08-OCT-1993;   93WO-NO00151.
XX
PR   08-OCT-1992;   92GB-0021163.
XX
PA   (NOBI-) NOBIPOL NOBIPOLS FORSKNINGSSTIFTELSE.
PA   (PRON-) PRONOVA BIOPOLYMER AS.
XX
PI   Ertesvag H,  Larsen B,  Valla S,  Skjak-braek G;
XX
DR   WPI; 1994-151310/18.
DR   P-PSDB; AAR54628, AAR54629, AAR54630, AAR54631.
XX
PT   New nucleic acid encoding mannuronan C-5-epimerase - and derived
PT   vectors and polypeptide, for prodn. of alginate of controlled G
PT   or M content, useful e.g. as immunostimulants or for cell
PT   immobilisation.
XX
PS   Claim 5; Page 34-45; 78pp; English.
XX
CC   The DNA can be used to produce the epimerase enzymes for the
CC   microbial production of alginates, esp. those of G (guluronic acid)
CC   block content 75-98% or those with a specific M (mannuronic acid)/G
CC   block content.  Inactivation of specific enzyme activities allows
CC   prodn. of pure poly-M alginates or those of G block content below
CC   25%.  Alginates of high M content are powerful, nontoxic
CC   immunostimulants while those of high G content can be used for cell
CC   immobilisation.
XX
SQ   Sequence 12588 BP; 2405 A; 4469 C; 3988 G; 1725 T; 1 other;

  Query Match             5.7%;  Score 57.8;  DB 15;  Length 12588;
  Best Local Similarity   43.3%;  Pred. No. 0.025;
  Matches  437;  Conservative    0;  Mismatches  557;  Indels   15;  Gaps    3;

Qy      20 TCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCG 79
           ||  |||| ||| |           ||||   |  | |||||  ||       ||    ||
Db    7651 TCTACACCACCACCGGCACCTGGATCGAAGGCAACACCATCGTCGGCTCGGCCAACTCCA 7710

Qy      80 GCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCG 139
           ||  |  ||||||  |  ||||| |||      |||  |  |||||  | ||||  | |||
Db    7711 CCTATGGCATCCAGGAGCGCGACGACGGCACCGACTACAGCAGCCTCTACGCCAACAGCG 7770

Qy     140 TCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACA 199
           ||  |    |  ||  ||  |   |||     | | |  || ||   |  | ||| ||||
Db    7771 TCAGCAATGTGCAGAACGGCTCGGTGCGCCTCTACGGCGCCAACTCCGTCGTCTCCGACC 7830

Qy     200 CCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCT 259
                 |           |  ||||  |   | | ||   |       |  | |||
Db    7831 TGCCCGGCACCGGCCAGCAGGCGACCCTCGAAGGCACGGCCG------GCAACGACACGC 7884

Qy     260 TCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACG 319
           |  |   | |||||  ||  ||   | ||||||    ||    |||    |||||| |||
Db    7885 TTGGCGGCAGCGACGCCCACGAGACGCTGCTCGGGCTGGACGGCAACGACCGCCTGAACG 7944

Qy     320 CCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGC 379
           |||   | |||| |||   || | ||    | ||  |      |      ||
Db    7945 GCGGCGCCGGCAACGACATCCTCGACGGCGGCGCCGGGCGCGACAACCTGACCGGCGGCG 8004

Qy     380 TGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGA 439
           | |||       |  |      |  |||    |   |    |        | |  |||  |
Db    8005 CGGGCGCCGACCTGTTCCGCGTCTCCGCGCGCACCGACAGCTACCGCACCGACAGCGCCA 8064

Qy     440 AGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCT 499
           |      |  | |||||  |    |    || | | |  |    ||||||  ||
Db    8065 GCTTCAACGACCTGATCACCGACTTCGACGCCAGCCAGGACCGCATCGACCTGTCCGCGC 8124

Qy     500 TCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCG 559
           |      | ||| || |  |  ||||   | ||  | | ||| ||  || ||| ||  ||
Db    8125 TGGGCTTCACCGGGCTGGGCGACGGCTATAACGGCACCCTGCTGCTGCAGGTCAGCGCCG 8184

Qy     560 CCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAG 619
           |||| ||  ||||       || | |  ||     |      |||    |    ||
Db    8185 ACGGCAGCCGCACCTATCTGAAGAGCCTGGAGGCGGATGCCGAGGGGCGGCGTTTCGAGA 8244

Qy     620 CCAGTTTCGAAGCCAAGGACCGTGGTGCGC---GCACCGCGATCCTGCTCGACTCGGACA 676
           |    | || | |||   |   ||   ||   |  |||  | |||||||  |  |  ||
Db    8245 TCGCCCTGGACGGCAACTTCGCCGGCCTGCTCGGTGCCGGCAACCTGCTCTTCGAGCGCA 8304

Qy     677 ACTGCGTGGCCGAAGGTCCGGGCTTCAACG------TGTGCATCGTCAAGGACGGCAAGC 730
           |  | | |  | | | | |||  |||||       |  | | |       | ||  |
Db    8305 CCGCCATCGAGGGGGATGCCGGCGACAACGCCCTGCTCGGTACCTCGGCCGCCGAGACAT 8364

Qy     731 TGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGG 790
           ||  |  ||   || | ||||    ||  | |  |                |  ||   | |
Db    8365 TGCTCGGCCACGCCGGCAACGACACGCTCGACGGCGGGGCCGGCGACGACATCCTGGTCG 8424

Qy     791 CCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACG 850
           || |     | ||  |     ||||    ||| ||   |   |       | || |||||
Db    8425 GCGGCGCCGGGCGCGACAGCCTCACCGGCGGCGCCGGAGCGGACGTGTTCCGCTTCGACG 8484

Qy     851 CCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATG 910
           |           |   | || |  ||     ||| | ||  | |  | | ||    | |  |
Db    8485 CGCTGTCCGACAGCCAGCGCAACTACGACATCGGCGACAACCAGGGCGACCGCATCGCCG 8544

Qy     911 GCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCT 970
           |   || ||||||  | | | ||| ||    | | || ||| | ||   |  | | |  |
Db    8545 ACTTCGCGGTGGGCGAAGACAAGCTCGACGTATCGGCGCTGGGCTTCACCGGGCTGGGCG 8604
```

```
Qy    971 GGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTG 1019
          |    |  | |   | | || | || || | |    ||   || |
Db   8605 ACGGCTACAACGGCACCCTCGCCCTGGTGCTCAACAGCGCCGGCGACCG 8653


RESULT 14
AAQ64201
ID   AAQ64201 standard; cDNA; 5392 BP.
XX
AC   AAQ64201;
XX
DT   18-NOV-1994  (first entry)
XX
DE   Sequence comprising the snaA, snaB and snaC gene cluster.
XX
KW   Antibiotic; streptogramin; snaA; snaB; snaC; biosynthesis; enzyme;
KW   biosynthetic pathway; Streptomyces pristinaespiralis; ds.
XX
OS   Streptomyces pristinaespiralis.
XX
PN   FR2696189-A.
XX
PD   01-APR-1994.
XX
PF   25-SEP-1992;   92FR-0011441.
XX
PR   25-SEP-1992;   92FR-0011441.
XX
PA   (RHON ) RHONE POULENC RORER SA.
XX
PI   Blanc V,  Blanche F,  Crouzet J,  Jacques N,  Lacroix P;
PI   Thibaut D,  Zagorec M;
XX
DR   WPI; 1994-128286/16.
XX
PT   DNA involved in streptogramin antibiotic biosynthesis - for
PT   prodn. or bio-conversion of streptogramin(s) or prodn. of
PT   streptogramin intermediates, derivs. or hybrid antibiotics
XX
PS   Disclosure; Page 44-47; 83pp; French.
XX
CC   This sequence comprises the snaA, snaB and snaC genes which are
CC   involved in the biosynthesis of streptogramins, antibiotics active
CC   against Gram-positive bacteria.  The identification of the sequences
CC   encoding the enzymes involved in the biosynthetic pathway means that
CC   they can be isolated and manipulated.  Mutant microorganisms in
CC   which a step in the streptogramin biosynthetic pathway is blocked
CC   can be cultured to produce streptogramin intermediates, which may
CC   later be converted to streptogramin derivatives.  Recombinant cells
CC   may also be used for the bioconversion of streptogramins from one
CC   form to another or for the production of hybrid antibiotics.
XX
SQ   Sequence 5392 BP; 811 A; 2161 C; 1671 G; 749 T; 0 other;

  Query Match             5.6%;  Score 57.4;  DB 15;  Length 5392;
  Best Local Similarity   42.1%;  Pred. No. 0.03;
  Matches  389;  Conservative    0;  Mismatches  531;  Indels    3;  Gaps    1;

Qy     16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75
          |||  || |  |  |  |||| ||  ||  || |||| | |   |  || || |    |
Db   3630 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 3689

Qy     76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135
            | |     ||||  |  |  |  ||    || |||| |||   ||| |||   |
Db   3690 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 3749

Qy    136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195
            || |     | ||   | |     |||| |   ||   | |     |  | |||    |
Db   3750 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 3809

Qy    196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255
          |  |||   |        ||       || || || ||||||     | ||  |  |
Db   3810 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 3869


Qy    256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
             |     | ||| | ||  |  |||  |    | || ||    | ||||   || |
Db   3870 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 3929

Qy    316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372
          | | ||    |  | ||| |    ||   | | |   |||| |  ||  |   ||| ||
Db   3930 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 3989

Qy    373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
            || ||||| |      | |||   | |     |      ||||     ||     ||||||
Db   3990 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 4049

Qy    433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
          | ||  |  | ||  | |  || | || |      |  |     | |||||  ||||| |
Db   4050 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 4109

Qy    493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
          |   |  ||||      |  | | | | ||| |  | || |         |   | |  ||
Db   4110 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCGACATC 4169

Qy    553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
            || || |  ||    ||||  ||||      |    |   |  | ||  |       |
Db   4170 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 4229

Qy    613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
          |   |||     ||    ||||  |      |    | |    | ||| || || ||
Db   4230 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 4289

Qy    673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
          |  | |  | |     ||         |   ||    |  | | |  | | ||||| |
Db   4290 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 4349

Qy    733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
          |  |    |       ||   || |  ||  |  | ||    || | | ||    || |
Db   4350 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 4409

Qy    793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
              | ||  |   |    || ||  |  ||   |  | |    |      ||| || |
Db   4410 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 4469

Qy    853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
          | || |    | ||   |  || || |  |||| |      || |     || |   |
Db   4470 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 4529

Qy    913 GAGGCCGTGGGCAACGGCGAGCC 935
            || ||       || |||| |
Db   4530 CCGGTCGGCCTGTTCGTCGAGAC 4552


RESULT 15
AAQ06844
ID   AAQ06844 standard; DNA; 2271 BP.
XX
AC   AAQ06844;
XX
DT   05-MAR-1991  (first entry)
XX
DE   Amylase gene from Streptomyces griseus IMRU 3570.
XX
KW   Secondary metabolism activation factor gene; amylase gene;
KW   heterologous protein expression; promoter; SAF; ss.
XX
OS   Streptomyces griseus.
XX
FH   Key             Location/Qualifiers
FT   CDS             318..2155
FT                   /*tag=  a
FT                   /product= amylase
FT                   /note= "including leader sequence"
XX
PN   WO9014426-A.
XX
```

PD 29-NOV-1990.
XX
PF 18-MAY-1990; 90WO-GB00781.
XX
PR 22-SEP-1989; 89US-0410706.
PR 19-MAY-1989; 89US-0354265.
XX
PA (ISTF) ARS APPL RES SYST.
XX
PI Ortega AD, Gil JA, Garcia TV, Martin JF;
XX
DR WPI; 1990-375993/50.
DR P-PSDB; AAR08263.
XX
PT Secondary metabolism activation factor gene - isolated from
PT Streptomyces and used to increase prodn. of extracellular enzymes
PT or heterologous proteins
XX
PS Disclosure; Fig 11; 61pp; English.
XX
CC This endogenous gene, with a suitable recognition site BstEII, is
CC used for fusion with a foreign gene. The amylase gene encodes
CC an extracellular enzyme ensuring secretion of the foreign
CC polypeptide or protein. The combination of the SAF gene (AAQ06843)
CC with the amylase gene on a plasmid does not cause increase prodn.
CC of amylase. Thus, it is preferred to place the foreign protein
CC operatively linked to the secretion signal sequence of an
CC endogenous extracellular enzyme which is controllable by SAF and,
CC pref. further operatively linked to the SAF promoter (in the absence
CC of the native promoter), into the chromosomal DNA as opposed to a
CC plasmid.
CC See also AAQ06843.
XX
SQ Sequence 2271 BP; 358 A; 946 C; 686 G; 281 T; 0 other;

Query Match 5.6%; Score 57.2; DB 11; Length 2271;
Best Local Similarity 44.4%; Pred. No. 0.032;
Matches 319; Conservative 0; Mismatches 393; Indels 6; Gaps 2;

```
Qy      18 CCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGC 77
           || ||   ||  |  |  ||| | || |||  | | |  || || |  || | |     |
Db     635 CCACGCGGCCGGCGTCAAGGTCGTCGCCGACTCGGTCATCAACCACATGGCCGCGGGTTC 694

Qy      78 CGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGG 137
           |||| | |  | |     |||||  ||| |   | ||  || |       ||     |||
Db     695 CGGCACCGGCACCGGCGGCAGCGCGTACCAGAAGTACGACTACCCGGGCATCTGGTCCGG 754

Qy     138 CGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGA 197
           || || |  |  |||  ||     |  |     |||  | |  ||   | ||| | || |
Db     755 CGCCGACATGGACGACTGCCGCAGCGAGATCAACGACTACGGCAACCGCGCCAACGTCCA 814

Qy     198 CACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACAT 257
           | |   | ||  |    ||  ||  || ||  || | |   ||    | ||| |  |
Db     815 GAAC---TGCGAACTGGTCGGCCTCGCCGACCTCGACACCGGTGAGCCGTACGTCCGCGA 871

Qy     258 CTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGA 317
           |  |     || | ||| |   |||| | |  |   ||| | |  |    | ||| | ||
Db     872 CCGCATCGCCGCCTACCTCAACGACCTGCTCTTGCTCGGCGTGGACGGCTTCCGCATCGA 931

Qy     318 CGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCA 377
           |||||   |   || |  ||     |||||  |||||   |     |    | |     |
Db     932 CGCCGCCAAGCACATGCCCGCC---GCCGACCTCACCGCCATCAAGGCCAAGGTCGGCAA 988

Qy     378 GCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGA 437
           | ||   |  |    | |   | || ||  | | |    ||       | ||  |
Db     989 CGGGAGCACGTACTGGAAGCAGGAGGCCATCCACGGCGCGGGCGAGGCCGTCCAGCCCAG 1048

Qy     438 GAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGC 497
           || ||||   ||    |  |   |||| || |  |||||||  |     |  | |||
Db    1049 CGAGTACCTCGGCACCGGCGACGTCCAGGAGTTCCGCTACGCCCGCGACCTCAAGCGGGT 1108

Qy     498 CTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCG 557
           ||||| |  ||   | | |  |  |   |  | |     || |    |      |
```

```
Db    1109 CTTCCAGAACGAGAACCTGGCCCACCTGAAGAACTTCGGCGAGGACTGGGGCTACATGGC 1168

Qy     558 CGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGC 617
           ||||     | |||||  |     || | |  ||  | || |||| |    ||||
Db    1169 GAGCGGCAAGTCCGCCGTCTTCGTCGACAACCACGACACCGAGCGGGGCGGCGACACCCT 1228

Qy     618 AGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAA 677
           |       | | |   | |           |  ||   || |||| | || |  | |
Db    1229 CAACTACAAGAACGGCTCCGCCTACACCCTCGCCGGCGTCTTCATGCTGGCCTGGCCCTA 1288

Qy     678 CTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCC 735
           | ||    | || |  |    |  | ||| || || || | | |||| |   | | ||
Db    1289 CGGCTCCCCGGACGTCCACTCCGGCTACGAGTTCACCGACCACGACGCCGGCCCGCCC 1346
```

Search completed: April 24, 2003, 19:58:40
Job time : 336 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 19:54:42 ; Search time 1541 Seconds
(without alignments)
10719.920 Million cell updates/sec

Title: US-10-067-291-2
Perfect score: 1020
Sequence: 1 atgactgctctttcagacct.....tcgaaacgatcgaatactga 1020

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	93.6	9.2	657	13	BI997754	BI997754 1031051G0
	2	92	9.0	662	17	BH190060	BH190060 ATXOA37TR
	3	82.4	8.1	587	17	BH190170	BH190170 ATXOD39TF
c	4	59.8	5.9	925	17	CNS0091P	AL053013 Drosophil
c	5	58.8	5.8	433	17	BH190123	BH190123 ATXOA70TF
	6	58	5.7	550	17	AZ935431	AZ935431 BJ__Ba000
	7	58	5.7	775	17	AZ933606	AZ933606 BJ__Ba000
	8	56.4	5.5	582	17	AZ934707	AZ934707 BJ__Ba000
	9	56.4	5.5	690	17	AZ933996	AZ933996 BJ__Ba000
	10	55.4	5.4	488	13	BI720731	BI720731 1031051G0
	11	54.6	5.4	478	10	AV633533	AV633533 AV633533
	12	54.6	5.4	574	10	AW757746	AW757746 874003D03
	13	54.6	5.4	607	14	BQ812089	BQ812089 1030027G0
	14	54.6	5.4	621	10	AW757472	AW757472 874001D07
	15	54.4	5.3	2299	11	AY106831	AY106831 Zea mays
c	16	54	5.3	613	10	AV927992	AV927992 AV927992
	17	54	5.3	640	10	AV923545	AV923545 AV923545
	18	54	5.3	642	10	AV923643	AV923643 AV923643
	19	54	5.3	653	10	AV923535	AV923535 AV923535
	20	54	5.3	655	10	AV923642	AV923642 AV923642
c	21	54	5.3	1529	17	AQ012108	AQ012108 8059P1A08
	22	53.6	5.3	533	10	AV644467	AV644467 AV644467
	23	53.6	5.3	572	14	BQ818594	BQ818594 1030071H0
	24	53.6	5.3	583	10	AW661017	AW661017 832009E04
c	25	53.4	5.2	502	10	AV947295	AV947295 AV947295
	26	53.4	5.2	590	12	BF474082	BF474082 WHE0840_H
	27	53.4	5.2	621	13	BJ277231	BJ277231 BJ277231
	28	53.2	5.2	625	13	BM441667	BM441667 EBed07_SQ
	29	53	5.2	585	10	AW758233	AW758233 874006G05
	30	53	5.2	636	10	AW721456	AW721456 833014A04
	31	52.6	5.2	519	10	AV644805	AV644805 AV644805
c	32	52.4	5.1	646	17	BH513273	BH513273 BOGUU01TF
	33	52.4	5.1	666	14	BQ767519	BQ767519 EBro08_SQ
	34	52.2	5.1	925	17	CNS0091P	AL053013 Drosophil
	35	52.2	5.1	1649	11	AY109876	AY109876 Zea mays
	36	52.2	5.1	3134	11	AY109500	AY109500 Zea mays
	37	52	5.1	554	14	BQ467985	BQ467985 HR01J03r
	38	52	5.1	752	12	BF261063	BF261063 HVSMEf002
	39	51.8	5.1	362	12	BG606125	BG606125 WHE2960_G
	40	51.8	5.1	646	14	BQ812088	BQ812088 1030027G0
	41	51.4	5.0	383	14	BQ467983	BQ467983 HR01J01r
c	42	51.4	5.0	546	14	BM737510	BM737510 952056F05
	43	51.2	5.0	555	10	AV390888	AV390888 AV390888
c	44	51	5.0	574	10	AW155392	AW155392 mgie0012C
	45	51	5.0	574	10	AW707791	AW707791 832012G12

ALIGNMENTS

RESULT 1
BI997754
LOCUS BI997754 657 bp mRNA linear EST 25-OCT-2001
DEFINITION 1031051G01.y2 C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI997754
VERSION BI997754.1 GI:16432528
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 657)
AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .657
/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoRI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 116 a 220 c 231 g 90 t
ORIGIN

Query Match 9.2%; Score 93.6; DB 13; Length 657;
Best Local Similarity 50.0%; Pred. No. 2.3e-08;
Matches 262; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

```
Qy    449 CCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCG 508
          ||||   ||||         | || | ||  ||| | | ||         |  |
Db    133 CCAACACCACCATCGGCAAGCCCACCATCGTAATCGTGCCCGAGTACAAGGAGGCCTCGC 192

Qy    509 CCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCA 568
          | | || ||      |   ||    | |  |  ||   || || || || | ||  | |
Db    193 CGGTGCCGAGGGAGCGGGGCATAAGGCTGTTCACGGTGCACGTGCGGCGTGGCGCGCCCG 252

Qy    569 ACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCG 628
          ||      |||||       ||   | ||      |     |   || ||| ||| |  ||
Db    253 ACGTGCAGGACCCCGGATGGAACTCCCACTCCAAGCTCAACTGCATCGCTGCCTGCATCC 312

Qy    629 AAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCG 688
          | |||||  |    |||| |  |   ||  |  |||| ||| || |   || |||||||
Db    313 AGGCCAACAAGGCGGGTGTGGACGAGGCTCTGATGCTGGACCCGCAGGGCTTCGTGGCCA 372

Qy    689 AAGGTCCGGGCTTCAACGTGTGCATCGT---CAAGGACGGCAAGCTGGCCTCCCCGTCCC 745
          |      ||  |||| | | ||||||   |||||  ||     |  |||  ||    | |
Db    373 CCTGCAACAGCACCAACTTCTTCATCGTGCGCAAGGGCGAGGTGTGGGCGCCCAGCCCGC 432

Qy    746 GGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCA 805
          || ||  | |||  ||||||||||    |  |||| ||   |||  | | |||   ||||
Db    433 GGCACCAGCTGCGCGGCATCACCCGGGCGCGGGTGCTCACGCTGTGCCAGCAGCACGGCA 492

Qy    806 TCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGG 865
          ||      |      |   ||| ||  |  |     |  | |||  |||||| |||    | |
Db    493 TCCCCTGCCGTGAGACGGACTTCTACCTCACGCAGGTGTACAGCGCCGAGGAGGCATTCG 552

Qy    866 CGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCA 925
          |      ||||    | ||| ||||  ||| |      | |||| |||  || | | |||
Db    553 TGACTGGCACCTTCGCCGGCGTCATCCCGGTGGTGGAGGTGGACGGCCGGGTCATAGGCG 612

Qy    926 ACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTC 969
          ||||  || ||| ||  || ||| |     | | | | | || |
Db    613 GCGGCAGGCGCGGCCCCGTGGCGGAGCGGCTGCAGCAGCTGTAC 656
```

RESULT 2
BH190060
LOCUS BH190060 662 bp DNA linear GSS 19-OCT-2001
DEFINITION ATXOA37TR ATXO Arabidopsis thaliana genomic clone ATXOA37, DNA sequence.
ACCESSION BH190060
VERSION BH190060.1 GI:16302842
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 662)
AUTHORS Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V., Feldblyum,T.V. and Fraser,C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC T24G4
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: ATXOA37TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1142. Caution: the DNA in this BAC may be from some non-Arabidopsis source
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .662
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="ATXOA37"
/clone_lib="ATXO"
/note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT 134 a 197 c 207 g 124 t
ORIGIN

Query Match 9.0%; Score 92; DB 17; Length 662;
Best Local Similarity 49.1%; Pred. No. 4.7e-08;
Matches 276; Conservative 0; Mismatches 280; Indels 6; Gaps 1;

```
Qy    449 CCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCG 508
          ||   ||    | ||| || |  ||| | |  |||| | |           | ||   ||
Db      2 CCTCTCTGTGCAATCGGGCGAGCATGTTCATCATCGCCTCGACCATCAGCCTCTATCCGA 61

Qy    509 CCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCA 568
            |||||             |   ||  |  |  |  |       ||||||| |||  | |
Db     62 AGGAGCACTATGAGAATGGCCTGGCCGTGATCACCTGTGCGACCCGCCGCCCCGCACCCG 121

Qy    569 ACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCG 628
          | |  |    |||     | |||        | |       |   || |   || |   | |
Db    122 CCGCGCTGATGCCGCAGGTGAAGTCGCTGAACTACCTGAACAACATCATGGCGAAGGTGG 181

Qy    629 AAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCG 688
          |||| |   |    |||||||      ||  |  ||||| ||  | |   || |||||| |
Db    182 AAGCGATCCAGGCCGGTGCGCTGGAAGCCGTGATGCTCAACGAGCAGGGCTACGTGGCGG 241

Qy    689 AAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGA 748
          || |  |||||  |||  | | | |  ||   ||   |  ||||    | |||     ||
Db    242 AATGCACGGGCGACAATATTTTCCTGATCCGCGATAACGTGCTGCATACGCCGCAGGTGA 301

Qy    749 ------ACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGG 802
                 |||| ||   ||||||||||    |  | |  ||||| |||| ||||  |||
Db    302 GCGACGGCGCGCTGGACGGCATCACCCGCCGGGTGATCGTCGAATTGGCGGACCAATTGG 361

Qy    803 GCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGA 862
          |  |              |||| |  | ||| ||  |||  |||| |  || |||||||
Db    362 GAGTGCCGATACTGGAGCGCCAGTTGACCCGCTATGACATCTATGTGGCGGACGAGTGCT 421

Qy    863 TGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGG 922
          |          |||||| |  | |||||  ||| ||       || ||  ||  |    | |
Db    422 TCCTCACGGGCACCGCCGCGGAGGTCATTCCGGTCGTGGAACTCGACCGCCGGAAGATCG 481

Qy    923 GCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGG 982
          || | ||| |||| ||||  | |||  |    ||   |||  |||  ||   ||   |
Db    482 GCGATGGCAAGCCGGGTCCCGTTACGAAGCGGTTCATCGACGCCTTCAAGGAATTGGCCG 541

Qy    983 ACGAGCCGGGCCCGCTGATCGA 1004
```

```
          |  ||||  | | ||| ||
Db    542 GCAGCACGGGTACTCCGATTGA 563
```

RESULT 3
BH190170
LOCUS BH190170 587 bp DNA linear GSS 19-OCT-2001
DEFINITION ATXOD39TF ATXO Arabidopsis thaliana genomic clone ATXOD39, DNA sequence.
ACCESSION BH190170
VERSION BH190170.1 GI:16303087
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 587)
AUTHORS Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V., Feldblyum,T.V. and Fraser,C.M.
TITLE Survey sequencing of Arabidopsis thaliana BAC T24G4
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: ATXOD39TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
From Wash. U contig 1142. Caution: the DNA in this BAC may be from some non-Arabidopsis source
Seq primer: TF
Class: sheared ends.

FEATURES Location/Qualifiers
source 1. .587
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="ATXOD39"
/clone_lib="ATXO"
/note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC DNA inserted into pHOS2 using BstXI linkers"

BASE COUNT 112 a 176 c 192 g 107 t
ORIGIN

Query Match 8.1%; Score 82.4; DB 17; Length 587;
Best Local Similarity 52.2%; Pred. No. 3.2e-06;
Matches 210; Conservative 0; Mismatches 186; Indels 6; Gaps 1;

```
Qy    552 CCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCT 611
          ||||||| |||  | |  | |  |    |||    | |||       | |      |
Db    186 CCGCCGCCCCGCACCCGCCGCGCTGATGCCGCAGGTGAAGTCGCTGAACTACCTGAACAA 245

Qy    612 CACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTC 671
          || |   || |   | ||||| |   |    |||||||     ||  |  ||||| ||
Db    246 CATCATGGCGAAGGTGGAAGCGATCCAGGCCGGTGCGCTGGAAGCCGTGATGCTCAACGA 305

Qy    672 GGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCT 731
          | |   || |||||| ||| |  |||||  |||  | | | |  ||   ||   |  |||
Db    306 GCAGGGCTACGTGGCGGAATGCACGGGCGACAATATTTTCCTGATCCGCGATAACGTGCT 365

Qy    732 GGCCTCCCCGTCCCGGAA------CGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGA 785
          |    | |||     ||       |||| ||   ||||||||||    |  | |  ||||
Db    366 GCATACGCCGCAGGTGAGCGACGGCGCGCTGGACGGCATCACCCGCCGGGTGATCGTCGA 425

Qy    786 ACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTA 845
          | |||| |||||  ||||  |             |||| |  | ||| ||  |||  ||||
Db    426 ATTGGCGGACCAATTGGGAGTGCCGATACTGGAGCGCCAGTTGACCCGCTATGACATCTA 485

Qy    846 CGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCT 905
          |  || |||||||   |          ||||||  |  | |||||  ||| ||      ||
Db    486 TGTGGCGGACGAGTGCTTCCTCACGGGCACCGCCGCGGAGGTCATTCCGGTCGTGGAACT 545

```
Qy    906 GGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGAC 947
          ||  ||  |    | ||| | ||| |||| |||||  | ||
Db    546 CGACCGCCGGAAGATCGGCGATGGCAAGCCGGGTCCCGTTAC 587
```

RESULT 4
CNS0091P/c
LOCUS CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match 5.9%; Score 59.8; DB 17; Length 925;
Best Local Similarity 14.6%; Pred. No. 0.077;
Matches 55; Conservative 180; Mismatches 136; Indels 6; Gaps 1;

```
Qy    527 CCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCA 586
          |: :|:|:  |: ::|::::: |: : ::: ::|::::::  : ::: :: :::| ::
Db    925 CSBSCSCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSS 866

Qy    587 TCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTG 646
          :  : ||  | |: : |  |  : : |: : :::     : : : ::|:|   : || :
Db    865 SSGTSSACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGS 806

Qy    647 CGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACG 706
          | :  :::::| :: :: |:|:  :  :  : ::: ::: | |||   :::   :  ::
Db    805 GGASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSS 746

Qy    707 TGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCA 766
          : |      :::::|::::   : :::| :: :: :|:::  :: :  :::::::| ||
Db    745 ASAG------SVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCA 692

Qy    767 CCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACG 826
          :|:    ::: : :|  ::   ::::: ::    :: :| ::| :::|::: ::|:: ::
Db    691 SCSCCCTSWSCSCSTSASMSAARSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSS 632
```
```

```
Qy    827 TCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGG 886
          :||  :::: |    : |:: ::::: :: |  :  || ||   :  ::: :|::||:|
Db    631 GSACGBSMSSGGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSG 572

Qy    887 TCACACCGATCAACTCG 903
          : : ::|  |: : ::
Db    571 SVCSCSSGCMCRCSCSS 555


RESULT 5
BH190123/c
LOCUS       BH190123                433 bp    DNA     linear   GSS 19-OCT-2001
DEFINITION  ATXOA70TF ATXO Arabidopsis thaliana genomic clone ATXOA70, DNA
            sequence.
ACCESSION   BH190123
VERSION     BH190123.1  GI:16302981
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T24G4
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: ATXOA70TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1142. Caution: the DNA in this BAC may be from
            some non-Arabidopsis source
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .433
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="ATXOA70"
                     /clone_lib="ATXO"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      100 a    135 c    122 g     76 t
ORIGIN

  Query Match             5.8%;  Score 58.8;  DB 17;  Length 433;
  Best Local Similarity   48.0%;  Pred. No. 0.11;
  Matches  168;  Conservative    0;  Mismatches  182;  Indels    0;  Gaps    0;

Qy    144 CTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGG 203
          |||| | || ||||| |   ||   |     |||||||||||||||| | |||||   |||
Db    358 CTGGTTGGACGGCGAGTTGGTGGATGAATCCGAAGCGAAGATCTCCGTGTTCGATCACGG 299

Qy    204 ATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCG 263
            |   | |   |||   |  || |   | |   |   | |   || ||    |||||||
Db    298 GCTGCTTTACGGCGACGGGGTCTTCGAGGGCATCCGCTTTTACAACCGCCGGGTCTTCCG 239

Qy    264 GCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGG 323
          || |  || ||| |   ||| ||| ||||    ||   | | |    |||||   |
Db    238 CCTGGAGGAGCACATCCGCCGTTTGTTCGATTCCGCCCGCGCGATCGTGCTGGAGATCCC 179

Qy    324 GTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCG 383
          |  |   | || ||  |   | || ||     | |   |||| |   |      ||
Db    178 TTGGACTCAAGAGGAGGTCTGCAAGTTCGTGGTGGAAACCGTGGCTGCGAACGGTCTCGA 119

Qy    384 CGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGA 443
          |||    |   |   |||  ||| || || || | |||    |         |   | |
```

```
Db    118 CGACGGCTATATCCGGCTGGTCGTGACGCGCGGTTCCGGCGGCCTGGGCTTGAACCCGTA 59

Qy    444 CCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGT 493
          |||| |||||   |     | |   ||||  | | | |     ||| |
Db     58 TCTGTGCAAGCGGGCGAGCATGTTCATCATCGCCTCGACCATCAGCCTCT 9


RESULT 6
AZ935431
LOCUS       AZ935431                550 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0003F10r B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ935431
VERSION     AZ935431.1  GI:13778155
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 509.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      101 a    180 c    158 g    111 t
ORIGIN

  Query Match             5.7%;  Score 58;  DB 17;  Length 550;
  Best Local Similarity   51.1%;  Pred. No. 0.16;
  Matches  136;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy    662 TGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGG 721
          | | ||||   | |  |     |||||  | | |   | ||||  |||  |  |||   |
Db    138 TTCGCGACCTCGTCGTCAAGCAGGCCGGCGCTTCCAACGTCAAGCTGTCCGACGTGGCGA 197

Qy    722 ACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGT 781
          |||  | ||||| ||||  |       || ||| |   ||| ||| | | |||   | |
Db    198 ACGTTACGCTGGGCTCCGAGGATTACGACTCGTCGGTCGGCTTCAACGGCAAGCGCGCGG 257

Qy    782 TCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAAC 841
          || |  | | |  |||| |||       | |||| ||| | |||||||| ||   | |  |
Db    258 TCTATATCGGCATCCAGGTGGCGCCAAACGCCAACCTTCTCGACGTCATCAAAGGCGTCC 317

Qy    842 TCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACT 901
          |  || |     |||  | | ||||        | | |||   | ' | |   |||| ||
Db    318 GCGCCGTCTATCCCGACATCAAGGCGCAGCAGCCGGAGGGGCTGAACTCCGAGATCATCT 377

Qy    902 CGCTGGATGGCGAGGCCGTGGGCAAC 927
                |||   ||||  || |
Db    378 ACGACTCCACCGATTTCGTGAACAGC 403


RESULT 7
AZ933606
```

```
LOCUS       AZ933606                775 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0001E09f B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ933606
VERSION     AZ933606.1  GI:13775666
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 716.
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      135 a    241 c    233 g    164 t      2 others
ORIGIN

  Query Match             5.7%;  Score 58;  DB 17;  Length 775;
  Best Local Similarity   51.1%;  Pred. No. 0.17;
  Matches  136;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy     662 TGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGG 721
           | | ||||   | |  |     |||||  | | |   | ||||  ||| |  |||   |
Db     139 TTCGCGACCTCGTCGTCAAGCAGGCCGGCGCTTCCAACGTCAAGCTGTCCGACGTGGCGA 198

Qy     722 ACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGT 781
           |||  | |||||  ||||  |        || ||| |   ||| ||| | | |||   | |
Db     199 ACGTTACGCTGGGCTCCGAGGATTACGACTCGTCGGTCGGCTTCAACGGCAAGCGCGCGG 258

Qy     782 TCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAAC 841
           || |  | | |  |||| |||      | |||| ||| | |||||||| ||   | | |
Db     259 TCTATATCGGCATCCAGGTGGCGCCAAACGCCAACCTTCTCGACGTCATCAAAGGCGTCC 318

Qy     842 TCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACT 901
           |  || |     |||  | | ||||       | | |||   |  | |   |||| ||
Db     319 GCGCCGTCTATCCCGACATCAAGGCGCAGCAGCCGGAGGGGCTGAACTCCGAGATCATCT 378

Qy     902 CGCTGGATGGCGAGGCCGTGGGCAAC 927
                  |||   ||||  || |
Db     379 ACGACTTCACCGATTTCGTGAACAGC 404


RESULT 8
AZ934707
LOCUS       AZ934707                582 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0002D18r B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ934707
VERSION     AZ934707.1  GI:13776767
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence start: 20
            High quality sequence stop: 548.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      114 a    186 c    159 g    123 t
ORIGIN

  Query Match             5.5%;  Score 56.4;  DB 17;  Length 582;
  Best Local Similarity   50.8%;  Pred. No. 0.32;
  Matches  135;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy     662 TGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGG 721
           | | ||||   | |  ||     ||||  | |     | ||||  ||| |  |||   |
Db     179 TTCGCGACCTCGTCGTCTAGCAAGCCGGCGCTTACAACGTCAAGCTGTCCGACGTGGCGA 238

Qy     722 ACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGT 781
           |||  | ||||| ||||  |        || ||| |   ||| ||| | | |||   | |
Db     239 ACGTTACGCTGGGCTCCGAGGATTACGACTCGTCGGTCGGCTTCAACGGCAAGCGCGCGG 298

Qy     782 TCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAAC 841
           || |  | | |  |||| |||      | |||| ||| | |||||||| ||   | | |
Db     299 TCTATATCGGCATCCAGGTGGCGCCAAACGCCAACCTTCTCGACGTCATCAAAGGCGTCC 358

Qy     842 TCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACT 901
           |  || |     |||  | | ||||       | | |||   |  | |   |||| ||
Db     359 GCGCCGTCTATCCCGACATCAAGGCGCAGCAGCCGGAGGGGCTGAACTCCGAGATCATCT 418

Qy     902 CGCTGGATGGCGAGGCCGTGGGCAAC 927
                  |||   ||||  || |
Db     419 ACGACTCCACCGATTTCGTGAACAGC 444


RESULT 9
AZ933996
LOCUS       AZ933996                690 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0001L03f B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ933996
VERSION     AZ933996.1  GI:13776056
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
```

```
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence start: 2
            High quality sequence stop: 651.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      125 a    226 c    193 g    145 t      1 others
ORIGIN
```

Query Match 5.5%; Score 56.4; DB 17; Length 690;
Best Local Similarity 50.8%; Pred. No. 0.33;
Matches 135; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
Qy     662 TGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGG 721
Db     139 TTCGCGACCTCGTCGTCAAGCAGGCCGGCGCTTCCAACGTCAAGCTGTCCGACGTGGCGA 198

Qy     722 ACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGT 781
Db     199 ACGTTACGCTGGGCTCCGAGGATTACGACTCGTCGGTACGCTTCAACGGCAAGCGCGCGG 258

Qy     782 TCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAAC 841
Db     259 TCTATATCGGCATCCAGGTGGCGCCAAACGCCAACCTTCTCGACGTCATCAAAGGCGTCC 318

Qy     842 TCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACT 901
Db     319 GCGCCGTCTATCCCGACATCAAGGCGCAGCAGCCGGAGGGGCTGAACTCCGAGATCATCT 378

Qy     902 CGCTGGATGGCGAGGCCGTGGGCAAC 927
Db     379 ACGACTCCACCGATTTCGTGAACAGC 404
```

```
RESULT 10
BI720731
LOCUS       BI720731                488 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  1031051G01.y1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI720731
VERSION     BI720731.1  GI:15696426
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      92 a    167 c    165 g     64 t
ORIGIN
```

Query Match 5.4%; Score 55.4; DB 13; Length 488;
Best Local Similarity 49.3%; Pred. No. 0.49;
Matches 174; Conservative 0; Mismatches 176; Indels 3; Gaps 1;

```
Qy     449 CCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCG 508
Db     133 CCAACACCACCATCGGCAAGCCCACCATCGTAATCGTGCCCGAGTACAAGGAGGCCTCGC 192

Qy     509 CCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCA 568
Db     193 CGGTGCCGAGGGAGCGGGGCATAAGGCTGTTCACGGTGCACGTGCGGCGTGGCGCGCCCG 252

Qy     569 ACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCG 628
Db     253 ACGTGCAGGACCCCGGATGGAACTCCCACTCCAAGCTCAACTGCATCGCTGCCTGCATCC 312

Qy     629 AAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCG 688
Db     313 AGGCCAACAAGGCGGGTGTGGACGAGGCTCTGATGCTGGACCCGCAGGGCTTCGTGGCCA 372

Qy     689 AAGGTCCGGGCTTCAACGTGTGCATCGT---CAAGGACGGCAAGCTGGCCTCCCCGTCCC 745
Db     373 CCTGCAACAGCACCAACTTCTTCATCGTGCGCAAGGGCGAGGTGTGGGCGCCCAGCCCGC 432

Qy     746 GGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAG 798
Db     433 GGCACCAGCTGCGCGGCATCACCCGGGCGCGGGTGCTCACGCTGTGCCAGCAG 485
```

```
RESULT 11
AV633533
LOCUS       AV633533                478 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV633533 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HC021h07_r 5', mRNA sequence.
ACCESSION   AV633533
VERSION     AV633533.1  GI:10776853
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
```

```
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HC021h07_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT       85 a    158 c    156 g     79 t
ORIGIN

  Query Match             5.4%;  Score 54.6;  DB 10;  Length 478;
  Best Local Similarity   47.0%;  Pred. No. 0.69;
  Matches  214;  Conservative    0;  Mismatches  229;  Indels   12;  Gaps    1;

Qy    554 GCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCA 613
          ||||| |  |||||  |  ||| |  |   ||  |||| || |  ||  |    |   |
Db     21 GCCGCTCTCGCCGCGCCGTCGTTGTGCGCGCCGGCAAGTACGATGAGGAGCTGATTAAGA 80

Qy    614 CCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGG 673
          ||||  | ||   | |   ||||||  ||| |||             ||||||  |      |
Db     81 CCGCTGGCACCGTTGCCTCCAAGGGCCGCGGT------------ATCCTGGCCATGGACG 128

Qy    674 ACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGG 733
          |   |  ||   ||   ||   | || |  ||     |  |||  || ||  | ||  |
Db    129 AGTCCAACGCCACCTGCGGCAAGCGCCTGGACTCCATCGGCGTGGAGAACACCGAGGAGA 188

Qy    734 CCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCG 793
          |  ||   ||     || | ||| ||  | | |||        ||      |  |  |||
Db    189 ACCGCCGCGCCTACCGCGAGCTGCTGGTGACCGCCCCCGGCCTGGGCCAGTACATCTCCG 248

Qy    794 ACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCG 853
          |   ||     |||| |  ||||||    |  ||||  |   |       |   || ||
Db    249 GCGCTATCCTGTTCGAGGAGACCCTGTATCAGTCCACCGCCTCCGGCAAGAAGTTCGTCG 308

Qy    854 ACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCG 913
          | | | ||| || |   | || || |  |||  |||      .|||    ||||    |
Db    309 ATGTGATGAAGGAGCAGAACATCGTGCCCGGCATCAAGGTCGACAAGGGCCTGGTGCCCC 368

Qy    914 AGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGG 973
          | ||   ||||||  |||  | |     ||    ||| |  || ||||  |  ||| |
Db    369 TGTCCAACACCAACGGTGAGAGCTGGTGCATGGGCCTGGACGGCCTGGACAAGCGCTGCG 428

Qy    974 CGCTGATGGACGAGCCGGGCCCGCTGATCGAAACG 1008
          |  |    || || | ||| | |   |||  | |
Db    429 CTGAGTACTACAAGGCCGGCGCTCGCTTCGCCAAG 463


RESULT 12
AW757746
LOCUS       AW757746                574 bp    mRNA    linear   EST 03-MAY-2000
DEFINITION  874003D03.y1 C. reinhardtii CC-1690, Lambda Zap II Chlamydomonas
            reinhardtii cDNA, mRNA sequence.
ACCESSION   AW757746
VERSION     AW757746.1  GI:7687098
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage."
BASE COUNT      105 a    192 c    177 g    100 t
ORIGIN

  Query Match             5.4%;  Score 54.6;  DB 10;  Length 574;
  Best Local Similarity   47.0%;  Pred. No. 0.71;
  Matches  214;  Conservative    0;  Mismatches  229;  Indels   12;  Gaps    1;

Qy    554 GCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCA 613
          |||||  |  |||||  |  ||| |  |   ||  |||| || |  ||  |    |   |
Db    114 GCCGCTCTCGCCGCGCCGTCGTTGTGCGCGCCGGCAAGTACGATGAGGAGCTGATTAAGA 173

Qy    614 CCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGG 673
          ||||  | ||   | |   ||||||  ||| |||             ||||||  |      |
Db    174 CCGCTGGCACCGTTGCCTCCAAGGGCCGCGGT------------ATCCTGGCCATGGACG 221

Qy    674 ACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGG 733
          |   |  ||   ||   ||   | || |  ||     |  |||  || ||  | ||  |
Db    222 AGTCCAACGCCACCTGCGGCAAGCGCCTGGACTCCATCGGCGTGGAGAACACCGAGGAGA 281

Qy    734 CCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCG 793
          |  ||   ||     || | ||| ||  | | |||        ||      |  |  |||
Db    282 ACCGCCGCGCCTACCGCGAGCTGCTGGTGACCGCCCCCGGCCTGGGCCAGTACATCTCCG 341

Qy    794 ACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCG 853
          |   ||     |||| |  ||||||    |  ||||  |   |       |   || ||
Db    342 GCGCTATCCTGTTCGAGGAGACCCTGTATCAGTCCACCGCCTCCGGCAAGAAGTTCGTCG 401

Qy    854 ACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCG 913
          | | | ||| || |   | || || |  |||  |||      .|||    ||||    |
Db    402 ATGTGATGAAGGAGCAGAACATCGTGCCCGGCATCAAGGTCGACAAGGGCCTGGTGCCCC 461

Qy    914 AGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGG 973
          | ||   ||||||  |||  | |     ||    ||| |  || ||||  |  ||| |
Db    462 TGTCCAACACCAACGGTGAGAGCTGGTGCATGGGCCTGGACGGCCTGGACAAGCGCTGCG 521

Qy    974 CGCTGATGGACGAGCCGGGCCCGCTGATCGAAACG 1008
          |  |    || || | ||| | |   |||  | |
Db    522 CTGAGTACTACAAGGCCGGCGCTCGCTTCGCCAAG 556


RESULT 13
BQ812089
LOCUS       BQ812089                607 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  1030027G07.x2 C. reinhardtii CC-1690, Deflagellation (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BQ812089
```

```
VERSION     BQ812089.1  GI:22055862
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1030
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Deflagellation
                     (normalized), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Deflagellation library, constructed by John Davies
                     and Jeffrey McDermott, combines cDNAs from CC-1690 cells
                     which had been re-synthesizing flagella for 15, 30 and 60
                     min after being deflagellated by pH shock. PolyA mRNA was
                     purified from each sample, pooled and cDNA synthesized.
                     The cDNA was directionally cloned into lambda Zap II
                     (Stratagene) in the EcoRI (5') and XhoRI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al., (1996) Genome Research 6: 791-806."
BASE COUNT      115 a    202 c    183 g    107 t
ORIGIN

  Query Match             5.4%;  Score 54.6;  DB 14;  Length 607;
  Best Local Similarity   47.0%;  Pred. No. 0.72;
  Matches  214;  Conservative    0;  Mismatches  229;  Indels   12;  Gaps    1;

Qy     554 GCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCA 613
           ||||| |  ||||| |  ||| |  |   ||  |||| || |  ||  |     |   |
Db     163 GCCGCTCTCGCCGCGCCGTCGTTGTGCGCGCCGGCAAGTACGATGAGGAGCTGATTAAGA 222

Qy     614 CCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGG 673
           ||||  |  ||   | |   ||||||  ||| |||             ||||||  |      |
Db     223 CCGCTGGCACCGTTGCCTCCAAGGGCCGCGGT-----------ATCCTGGCCATGGACG 270

Qy     674 ACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGG 733
           |   |  ||   ||   ||   | || |  ||      |  |||  || ||  | ||  |
Db     271 AGTCCAACGCCACCTGCGGCAAGCGCCTGGACTCCATCGGCGTGGAGAACACCGAGGAGA 330

Qy     734 CCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCG 793
           |  ||   ||     || | ||| ||  | | |||        ||       |  |  |||
Db     331 ACCGCCGCGCCTACCGCGAGCTGCTGGTGACCGCCCCCGGCCTGGGCCAGTACATCTCCG 390

Qy     794 ACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCG 853
           |   ||      |||| |  ||||||    |   ||||  |   |       |   || ||
Db     391 GCGCTATCCTGTTCGAGGAGACCCTGTATCAGTCCACCGCCTCCGGCAAGAAGTTCGTCG 450

Qy     854 ACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCG 913
           | | | ||| ||  |   | || || |  |||  |||        |||    ||||    |
Db     451 ATGTGATGAAGGAGCAGAACATCGTGCCCGGCATCAAGGTCGACAAGGGCCTGGTGCCCC 510

Qy     914 AGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGG 973
           | ||     |||||| |||  | |     ||    |||  |  || ||||  |  ||| |
Db     511 TGTCCAACACCAACGGTGAGAGCTGGTGCATGGGCCTGGACGGCCTGGACAAGCGCTGCG 570

Qy     974 CGCTGATGGACGAGCCGGGCCCGCTGATCGAAACG 1008
           |   |    || || | ||| | |   |||  | |
Db     571 CTGAGTACTACAAGGCCGGCGCTCGCTTCGCCAAG 605


RESULT 14
AW757472
LOCUS       AW757472                621 bp    mRNA    linear   EST 03-MAY-2000
DEFINITION  874001D07.y1 C. reinhardtii CC-1690, Lambda Zap II Chlamydomonas
            reinhardtii cDNA, mRNA sequence.
ACCESSION   AW757472
VERSION     AW757472.1  GI:7686824
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage."
BASE COUNT      111 a    215 c    188 g    107 t
ORIGIN

  Query Match             5.4%;  Score 54.6;  DB 10;  Length 621;
  Best Local Similarity   47.0%;  Pred. No. 0.72;
  Matches  214;  Conservative    0;  Mismatches  229;  Indels   12;  Gaps    1;

Qy     554 GCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCA 613
           ||||| |  ||||| |  ||| |  |   ||  |||| || |  ||  |     |   |
Db     115 GCCGCTCTCGCCGCGCCGTCGTTGTGCGCGCCGGCAAGTACGATGAGGAGCTGATTAAGA 174

Qy     614 CCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGG 673
           ||||  |  ||   | |   ||||||  ||| |||             ||||||  |      |
Db     175 CCGCTGGCACCGTTGCCTCCAAGGGCCGCGGT-----------ATCCTGGCCATGGACG 222

Qy     674 ACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGG 733
           |   |  ||   ||   ||   | || |  ||      |  |||  || ||  | ||  |
Db     223 AGTCCAACGCCACCTGCGGCAAGCGCCTGGACTCCATCGGCGTGGAGAACACCGAGGAGA 282

Qy     734 CCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCG 793
           |  ||   ||     || | ||| ||  | | |||        ||       |  |  |||
Db     283 ACCGCCGCGCCTACCGCGAGCTGCTGGTGACCGCCCCCGGCCTGGGCCAGTACATCTCCG 342
```

```
Qy    794 ACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCG 853
          |   ||     |||| |  ||||||    |   ||||  |   |       |   || ||
Db    343 GCGCTATCCTGTTCGAGGAGACCCTGTATCAGTCCACCGCCTCCGGCAAGAAGTTCGTCG 402

Qy    854 ACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCG 913
          | | | ||| || |   | || || |  |||  |||       |||    ||||    |
Db    403 ATGTGATGAAGGAGCAGAACATCGTGCCCGGCATCAAGGTCGACAAGGGCCTGGTGCCCC 462

Qy    914 AGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGG 973
          | ||     |||||| ||| | |      ||    ||| |  || ||||  |  ||| |
Db    463 TGTCCAACACCAACGGTGAGAGCTGGTGCATGGGCCTGGACGGCCTGGACAAGCGCTGCG 522

Qy    974 CGCTGATGGACGAGCCGGGCCCGCTGATCGAAACG 1008
          |   |    || || | ||| | |   |||  | |
Db    523 CTGAGTACTACAAGGCCGGCGCTCGCTTCGCCAAG 557


RESULT 15
AY106831
LOCUS       AY106831               2299 bp    mRNA    linear   HTC 25-MAY-2002
DEFINITION  Zea mays PCO142079 mRNA sequence.
ACCESSION   AY106831
VERSION     AY106831.1  GI:21209909
KEYWORDS    HTC.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2299)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2299)
  AUTHORS   Coe,E.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1. .2299
                     /organism="Zea mays"
                     /db_xref="MaizeDB:638374"
                     /db_xref="taxon:4577"
                     /clone="PCO142079"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      475 a    790 c    658 g    376 t
ORIGIN

  Query Match             5.3%;  Score 54.4;  DB 11;  Length 2299;
  Best Local Similarity   52.1%;  Pred. No. 0.98;
  Matches  173;  Conservative    0;  Mismatches  151;  Indels    8;  Gaps    2;

Qy    651 CACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTG 710
          ||||    ||| |  | |||| | || | || | |          || | ||| |  |
Db    319 CACCCTCCTCCAGGGCTACTCCGGCATCCGCTTCGAGATCCTCGAGGCCATCACCAAGCT 378

Qy    711 CATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGG--CATCACC 768
          ||||  ||  | || ||  |  |||| ||  | | ||    | |  ||| |     |  |
Db    379 CATCAACACCGGCGTCAGCCGTGCCTGCCGCTGCGGGGCACCATCACCGCGTCGGGCGAC 438

Qy    769 CGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTC 828
          |      || ||| | |  | |||| ||  ||           ||| |  ||  | ||||
Db    439 CTCGTCCCGCTGTCCTACATCGCCGGCCTCATCACGGGGCGTCCCAACGCGCAGGCCGTC 498

Qy    829 AC------CAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGC 882
          ||      | || |  |  |  ||| |||||| | ||| | |||||  |  || || ||||
Db    499 ACGGTTGACGGCAGGAAGGTGGACGCCGCCGAGGCGTTCAAGGCGGCCGGCATCGAGGGC 558

Qy    883 GGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCA 942
          ||  ||     | ||||| |   ||| |||  |||| | | |||||||  | |||
Db    559 GGCTTCTTCAAGCTCAACCCCAAGGAGGGCCTGGCCATCGTCAACGGCACGTCCGTGGGG 618

Qy    943 CTGACGGTGGCCATCCGGGACCGGTTCTGGGC 974
          | || | |||  |  ||   | |||   ||
Db    619 TCGGCGCTCGCCGCCATGGTGTGCTTCGACGC 650


Search completed: April 24, 2003, 20:34:20
Job time : 1553 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 19:43:06 ; Search time 98 Seconds
(without alignments)
3191.942 Million cell updates/sec

Title: US-10-067-291-2
Perfect score: 1020
Sequence: 1 atgactgctctttcagacct.....tcgaaacgatcgaatactga 1020

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1020	100.0	1020	4	US-09-527-522-2	Sequence 2, Appli
2	66.2	6.5	15872	4	US-09-105-537-1	Sequence 1, Appli
3	57.8	5.7	12588	2	US-08-387-942C-1	Sequence 1, Appli
4	57.4	5.6	1208	2	US-08-403-852D-4	Sequence 4, Appli
5	57.4	5.6	1208	3	US-08-510-646B-4	Sequence 4, Appli
6	57.4	5.6	1208	4	US-09-231-818-4	Sequence 4, Appli
7	57.4	5.6	5392	2	US-08-403-852D-1	Sequence 1, Appli
8	57.4	5.6	5392	3	US-08-510-646B-1	Sequence 1, Appli
9	57.4	5.6	5392	4	US-09-231-818-1	Sequence 1, Appli
10	57	5.6	1140	3	US-09-023-173-4	Sequence 4, Appli
11	55.4	5.4	1162	4	US-09-173-300-21	Sequence 21, Appl
12	53.8	5.3	1294	3	US-09-025-691-2	Sequence 2, Appli
13	53.6	5.3	2634	1	US-08-196-218-31	Sequence 31, Appl
14	53.6	5.3	2634	1	US-08-681-953-31	Sequence 31, Appl
15	53.2	5.2	735	3	US-09-003-287-7	Sequence 7, Appli
16	51.6	5.1	3756	2	US-08-576-626A-1	Sequence 1, Appli
17	50.4	4.9	1086	4	US-09-173-300-14	Sequence 14, Appl
18	50.2	4.9	1835	4	US-09-417-704-2	Sequence 2, Appli
19	50	4.9	20235	1	US-07-642-734C-3	Sequence 3, Appli
20	50	4.9	20235	3	US-08-439-009A-3	Sequence 3, Appli
c 21	50	4.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 22	48.8	4.8	1693	3	US-09-320-878-23	Sequence 23, Appl
23	48.2	4.7	1473	2	US-08-541-033A-25	Sequence 25, Appl
24	48.2	4.7	1473	2	US-08-828-451-25	Sequence 25, Appl
25	48.2	4.7	1506	2	US-08-541-033A-23	Sequence 23, Appl
26	48.2	4.7	1506	2	US-08-828-451-23	Sequence 23, Appl
27	48.2	4.7	1969	2	US-08-541-033A-7	Sequence 7, Appli
28	48.2	4.7	1969	2	US-08-828-451-7	Sequence 7, Appli
29	48.2	4.7	2096	2	US-08-541-033A-19	Sequence 19, Appl
30	48.2	4.7	2096	2	US-08-828-451-19	Sequence 19, Appl
31	48.2	4.7	2099	2	US-08-541-033A-3	Sequence 3, Appli
32	48.2	4.7	2099	2	US-08-828-451-3	Sequence 3, Appli
33	48.2	4.7	2137	2	US-08-541-033A-18	Sequence 18, Appl
34	48.2	4.7	2137	2	US-08-828-451-18	Sequence 18, Appl
35	48.2	4.7	2140	2	US-08-541-033A-1	Sequence 1, Appli
36	48.2	4.7	2140	2	US-08-828-451-1	Sequence 1, Appli
37	48	4.7	1143	2	US-08-387-942C-19	Sequence 19, Appl
38	47.8	4.7	2064	1	US-08-343-428-1	Sequence 1, Appli
39	47.8	4.7	2167	2	US-08-461-775-9	Sequence 9, Appli
40	47.8	4.7	2167	3	US-09-031-606-9	Sequence 9, Appli
41	47.8	4.7	2668	2	US-08-461-775-11	Sequence 11, Appl
42	47.8	4.7	2668	3	US-09-031-606-11	Sequence 11, Appl
43	47.4	4.6	1288	1	US-08-440-856A-9	Sequence 9, Appli
44	47.2	4.6	8051	2	US-08-576-626A-2	Sequence 2, Appli
45	47	4.6	390	4	US-09-197-649-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-527-522-2
; Sequence 2, Application US/09527522
; Patent No. 6413752
; GENERAL INFORMATION:
; APPLICANT: Takashima, Yoshiki
; APPLICANT: Mitsuda, Satoshi
; TITLE OF INVENTION: Protein capable of catalyzing transamination
; TITLE OF INVENTION: stereoselectively, gene encoding said protein
; TITLE OF INVENTION: and use thereof
; FILE REFERENCE: Q58251
; CURRENT APPLICATION NUMBER: US/09/527,522
; CURRENT FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: JP 11/075511
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 11/088634
; PRIOR FILING DATE: 1999-03-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Mycobacterium aurum
; FEATURE:
; OTHER INFORMATION: SC-S423
; NAME/KEY: CDS
; LOCATION: (1)..(1017)
US-09-527-522-2

Query Match 100.0%; Score 1020; DB 4; Length 1020;
Best Local Similarity 100.0%; Pred. No. 7.6e-214;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60

Qy      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120

Qy     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180

Qy     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240

Qy     241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300
```

```
Db    241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300

Qy    301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360

Db    301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360

Qy    361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420

Db    361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420

Qy    421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480

Db    421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480

Qy    481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540

Db    481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540

Qy    541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600

Db    541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600

Qy    601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660

Db    601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660

Qy    661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720

Db    661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720

Qy    721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780

Db    721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780

Qy    781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840

Db    781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840

Qy    841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900

Db    841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900

Qy    901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960

Db    901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960

Qy    961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020

Db    961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020
```

RESULT 2
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match 6.5%; Score 66.2; DB 4; Length 15872;
Best Local Similarity 44.4%; Pred. No. 7.8e-06;
Matches 358; Conservative 0; Mismatches 443; Indels 6; Gaps 2;

```
Qy     50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109

Db    765 CCCGCGCCCTCGCCGACGGCGACCGTGTCCACGGCGTCATCCGCGCCAGCGCCGTCAACA 824

Qy    110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169

Db    825 ACGACGGAGCCACCCCGGGTCTCACCGTGCCCAGCAGGGCCGCCCAGGAGAAGGTGCTGC 884

Qy    170 CCGAAG---AAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCT 226

Db    885 GCGAGGCGTACCGGAAGGCGGCCCTGGACCCGTCCGCCGTCCAGTACGTCGAACTCCACG 944

Qy    227 ACACCGTCGCGCATGTATGGCACGGCA---ACATCTTCCGGCTCGGCGACCACCTGGACC 283

Db    945 GCACCGGAACCCCCGTCGGCGACCCCATCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCT 1004

Qy    284 GGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGG 343

Db   1005 CGGCGCGCCCCGCGGACGAACCCCTGCTCGTCGGCTCGGCCAAGACGAACGTCGGGCACC 1064

Qy    344 CCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGA 403

Db   1065 TCGAAGGCGCCGCCGGCATCGTCGGCCTCATCAAGACGCTCCTCGCGCTCGGCCGGCGCC 1124

Qy    404 CCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATC 463

Db   1125 GGATCCCGGCGAGCCTCAACTTCCGTACGCCCCACCCGGACATCCCGCTCGACACCCTCG 1184

Qy    464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523

Db   1185 GGCTCGACGTGCCCGACGGCCTGCGGGAGTGGCCGCACCCGGACCGCGAACTCCTCGCCG 1244

Qy    524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583

Db   1245 GCGTCAGCTCGTTCGGCATGGGCGGCACCAACGCCCACGTCGTCCTCAGCGAAGGCCCCG 1304

Qy    584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643

Db   1305 CCCAGGGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCCGTCGACAGCGGGGCCGCAC 1364

Qy    644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703

Db   1365 TGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGCGCCCAGGCCCGGCGCCTGC 1424

Qy    704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763

Db   1425 ACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCACTCGCCCGGTCGCTGGTCA 1484

Qy    764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823

Db   1485 CCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCC 1544

Qy    824 ACGTCACCAGCCGTGAACTCTACGACG 850

Db   1545 TCGACGGCCTCGGCGCCCTCGCCGCCG 1571
```

RESULT 3
US-08-387-942C-1
; Sequence 1, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH, LLP
;      STREET:  P.O.BOX 747
;      CITY:  FALLS CHURCH
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/387,942C
;      FILING DATE:  09-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR, GERALD M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1809-106P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12588 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Azotobacter vinelandii
;      STRAIN:  E
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  290..1951
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2227..6438
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  6702..9695
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9973..12588
US-08-387-942C-1

  Query Match             5.7%;  Score 57.8;  DB 2;  Length 12588;
  Best Local Similarity   43.3%;  Pred. No. 0.00051;
  Matches  437;  Conservative    0;  Mismatches  557;  Indels   15;  Gaps    3;

Qy      20 TCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCG 79
           ||  |||| ||| |          ||||   |  | |||||  ||      ||    ||
Db    7651 TCTACACCACCACCGGCACCTGGATCGAAGGCAACACCATCGTCGGCTCGGCCAACTCCA 7710

Qy      80 GCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCG 139
           ||  |  ||||||  |  ||||| |||     |||  |  |||||  | ||||  | |||
Db    7711 CCTATGGCATCCAGGAGCGCGACGACGGCACCGACTACAGCAGCCTCTACGCCAACAGCG 7770

Qy     140 TCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACA 199
           ||  |     |  ||  ||   |   |||     | | |  || ||    |  | ||| ||||
Db    7771 TCAGCAATGTGCAGAACGGCTCGGTGCGCCTCTACGGCGCCAACTCCGTCGTCTCCGACC 7830

Qy     200 CCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCT 259
               |          |  ||||  |   | | ||   |       |  | |||
Db    7831 TGCCCGGCACCGGCCAGCAGGCGACCCTCGAAGGCACGGCCG------GCAACGACACGC 7884

Qy     260 TCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACG 319
           |  |   | |||||  ||  ||   | ||||||    ||    |||    |||||| |||
Db    7885 TTGGCGGCAGCGACGCCCACGAGACGCTGCTCGGGCTGGACGGCAACGACCGCCTGAACG 7944

Qy     320 CCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGC 379
             |||   | |||| |||   || | ||    | ||  |     |     ||

Db    7945 GCGGCGCCGGCAACGACATCCTCGACGGCGGCGCCGGGCGCGACAACCTGACCGGCGGCG 8004

Qy     380 TGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGA 439
             | |||       |  |      |  |||     |   |     |          | |  |||  |
Db    8005 CGGGCGCCGACCTGTTCCGCGTCTCCGCGCGCACCGACAGCTACCGCACCGACAGCGCCA 8064

Qy     440 AGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCT 499
               |    |  | |||||  |     |    || | | | |   |    ||||||  ||
Db    8065 GCTTCAACGACCTGATCACCGACTTCGACGCCAGCCAGGACCGCATCGACCTGTCCGCGC 8124

Qy     500 TCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCG 559
           |      | ||| || |  |  ||||   | ||  | | ||| ||  || ||| ||  ||
Db    8125 TGGGCTTCACCGGGCTGGGCGACGGCTATAACGGCACCCTGCTGCTGCAGGTCAGCGCCG 8184

Qy     560 CCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAG 619
             |||| ||  ||||         || | |  ||     |       |||    |    ||
Db    8185 ACGGCAGCCGCACCTATCTGAAGAGCCTGGAGGCGGATGCCGAGGGGCGGCGTTTCGAGA 8244

Qy     620 CCAGTTTCGAAGCCAAGGACCGTGGTGCGC---GCACCGCGATCCTGCTCGACTCGGACA 676
             |   | || | |||   |   ||   ||   |  |||  | |||||||  |  |  ||
Db    8245 TCGCCCTGGACGGCAACTTCGCCGGCCTGCTCGGTGCCGGCAACCTGCTCTTCGAGCGCA 8304

Qy     677 ACTGCGTGGCCGAAGGTCCGGGCTTCAACG------TGTGCATCGTCAAGGACGGCAAGC 730
             | | |  |  | | | |||  ||||N       |  | | |       | ||  |
Db    8305 CCGCCATCGAGGGGGATGCCGGCGACAACGCCCTGCTCGGTACCTCGGCCGCCGAGACAT 8364

Qy     731 TGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGG 790
             ||  |  ||   || | ||||    ||  | | |  |                |  ||    | |
Db    8365 TGCTCGGCCACGCCGGCAACGACACGCTCGACGGCGGGGCCGGCGACGACATCCTGGTCG 8424

Qy     791 CCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACG 850
             || |     | ||  |      ||||    ||| ||   |   |       | || |||||
Db    8425 GCGGCGCCGGGCGCGACAGCCTCACCGGCGGCGCCGGAGCGGACGTGTTCCGCTTCGACG 8484

Qy     851 CCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATG 910
             |          |   |  || |  ||    ||| | ||  | | |  | ||    | |  |
Db    8485 CGCTGTCCGACAGCCAGCGCAACTACGACATCGGCGACAACCAGGGCGACCGCATCGCCG 8544

Qy     911 GCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCT 970
             |   || ||||||  | | | ||| ||     |  | || ||| | ||   |  | |  |
Db    8545 ACTTCGCGGTGGGCGAAGACAAGCTCGACGTATCGGCGCTGGGCTTCACCGGGCTGGGCG 8604

Qy     971 GGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTG 1019
             |     |  | |   |  | || | || || | |    ||   || |
Db    8605 ACGGCTACAACGGCACCCTCGCCCTGGTGCTCAACAGCGCCGGCGACCG 8653


RESULT 4
US-08-403-852D-4
; Sequence 4, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/403,852D
;       FILING DATE:  10-MAY-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/FR 93/00923
;       FILING DATE:  25-SEP-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 92/11441
;       FILING DATE:  25-SEP-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meyers, Kenneth J.
;       REGISTRATION NUMBER:  25,146
;       REFERENCE/DOCKET NUMBER:  03806.0054-00000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 408-4000
;       TELEFAX:  (202) 408-4400
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1208 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  S.pristinaespiralis
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1208
US-08-403-852D-4

  Query Match             5.6%;  Score 57.4;  DB 2;  Length 1208;
  Best Local Similarity   42.1%;  Pred. No. 0.0004;
  Matches  389;  Conservative    0;  Mismatches  531;  Indels    3;  Gaps    1;

Qy      16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75
           |||  || |  |  |  |||| ||  ||  || |||| | |   |  || || |    |
Db      73 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 132

Qy      76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135
             | |      ||||  |  |  |  ||    || |||| |||   ||| |||   |
Db     133 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 192

Qy     136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195
           || |     | ||   | |     |||| |   ||   | |     |  | |||    |
Db     193 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 252

Qy     196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255
             |  |||   |       ||       || || || ||||||      | ||  |  |
Db     253 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 312

Qy     256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
             |     | ||| | ||  |  |||  |    | || ||    | ||||   || |
Db     313 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 372

Qy     316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372
           | | ||    |  | ||| |    ||   | | |   |||| |  ||  |   ||| ||
Db     373 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 432

Qy     373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
             || ||||| |     | |||   | |    |     ||||     ||     ||||||
Db     433 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 492

Qy     433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
           | ||  |  | ||  | |  || | ||        |  |     | |||||  ||||| |
Db     493 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 552

Qy     493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
```

```
           |   |  ||||     |  | | | | ||| |  | || |        |   | |  ||
Db     553 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCGACATC 612

Qy     553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
             || || |  ||    ||||  ||||     |    |  | | ||   |        |
Db     613 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 672

Qy     613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
           |    |||      ||    ||||  |       |    | |    | ||| || || ||
Db     673 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 732

Qy     673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
           |  | |  | |      ||         |   ||     |  | | |  | | ||||| |
Db     733 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 792

Qy     733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
           |  |    |      ||   || |  ||   |  | ||    || | | ||    || |
Db     793 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 852

Qy     793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
             | ||  |  |    || ||  |  ||   |  | |    |      ||| || |
Db     853 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 912

Qy     853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
           | || |    | ||   |  | || || |  |||| |       || |    || |
Db     913 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 972

Qy     913 GAGGCCGTGGGCAACGGCGAGCC 935
             || ||      || |||| |
Db     973 CCGGTCGGCCTGTTCGTCGAGAC 995
```

```
RESULT 5
US-08-510-646B-4
; Sequence 4, Application US/08510646B
; Patent No. 6077699
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/510,646B
;      FILING DATE:  03-AUG-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,852
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-01000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1208 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1208
US-08-510-646B-4

  Query Match             5.6%;  Score 57.4;  DB 3;  Length 1208;
  Best Local Similarity   42.1%;  Pred. No. 0.0004;
  Matches  389;  Conservative    0;  Mismatches  531;  Indels    3;  Gaps    1;

Qy      16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75
           |||  || |  |  |  |||| ||  ||  || |||| | |   |  || || |    |
Db      73 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 132

Qy      76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135
             | |      ||||  |  |  |  ||    || |||| |||   ||| |||   |
Db     133 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 192

Qy     136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195
            || |     | ||   | |     |||| |   ||   | |     |  | |||    |
Db     193 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 252

Qy     196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255
           |  |||  |         ||       || || || ||||||      | ||  |  |
Db     253 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 312

Qy     256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
              |     | ||| | ||  |  |||  |    | || ||    | ||||   || |
Db     313 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 372

Qy     316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372
           | | ||    |  | ||| |    ||   | | |   |||| |  ||  |   ||| ||
Db     373 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 432

Qy     373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
           || ||||| |     | |||   | |    |     ||||     ||     ||||||
Db     433 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 492

Qy     433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
           | ||  |  | ||  | |  || | || |       |  |     | |||||  ||||| |
Db     493 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 552

Qy     493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
           |   |  ||||     |  | | | | ||| |  | || |         |   | |  ||
Db     553 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCGACATC 612

Qy     553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
           || || |  ||    ||||  ||||      |     |   | ||   |        |
Db     613 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 672

Qy     613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
           |    |||     ||    ||||  |       |    | |    | ||| || || ||
Db     673 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 732

Qy     673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
           |  | |  | |     ||        |   ||    |  | | |  | | |||||  |
Db     733 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 792

Qy     733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
             |  |    |      ||   || |  ||  |  | ||    || | | ||    || |
Db     793 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 852

Qy     793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
             | || |  |    || ||  |  ||   |   | |     |       ||| || |
Db     853 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 912

Qy     853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
           | || |    | ||   |  |  || || |  |||| |      || |     ||  |
Db     913 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 972

Qy     913 GAGGCCGTGGGCAACGGCGAGCC 935
             || ||       || |||| |
Db     973 CCGGTCGGCCTGTTCGTCGAGAC 995


RESULT 6
US-09-231-818-4
; Sequence 4, Application US/09231818
; Patent No. 6171846
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/231,818
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852
;      FILING DATE:  10-MAY-1995
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1208 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  S.pristinaespiralis
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1208
US-09-231-818-4

  Query Match             5.6%;  Score 57.4;  DB 4;  Length 1208;
  Best Local Similarity   42.1%;  Pred. No. 0.0004;
  Matches  389;  Conservative    0;  Mismatches  531;  Indels    3;  Gaps    1;

Qy      16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75
           |||  || |  |  |  |||| ||  ||  || |||| | |   |  || || |    |
Db      73 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 132

Qy      76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135
             | |      ||||  |  |  |  ||    || |||| |||   ||| |||   |
Db     133 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 192

Qy     136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195
           || |     | ||   | |     |||| |   ||   | |     |  | |||    |
Db     193 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 252

Qy     196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255
           |  |||   |        ||      || || || ||||||      | ||  |  |
Db     253 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 312

Qy     256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
             |    | ||| | ||  |  |||  |    | || ||    | ||||   || |
Db     313 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 372

Qy     316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372
           | | ||    |  | |||  |    ||   | | |   ||||  |  ||  |   ||| ||
Db     373 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 432

Qy     373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
           || |||||  |     | |||   | |     |     ||||    ||    ||||||
Db     433 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 492

Qy     433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
           | ||  |  | ||  | |  || | ||  |       |  |     | |||||  ||||| |
Db     493 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 552

Qy     493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
           |   | ||||     | | | | | ||| |  | ||  |         |   | |  ||
Db     553 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCGACATC 612

Qy     553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
           || || |  ||    ||||  ||||     |   |   |  | ||  |       |
Db     613 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 672

Qy     613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
           |   |||     ||    ||||  |       |   | |     | ||| || || ||
Db     673 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 732

Qy     673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
           |  | |  | |     ||         |   ||     |  | | |   | | |||| |
Db     733 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 792

Qy     733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
           |  |     |       ||   || |  ||  |  | ||    || | | ||    || |
Db     793 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 852

Qy     793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
             | || |  |    || ||  |  ||   |  | |    |       ||| || |
Db     853 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 912

Qy     853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
```

```
           | || |    | ||   |  |  || || |  |||| |      || |     || |
Db     913 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 972

Qy     913 GAGGCCGTGGGCAACGGCGAGCC 935
             || ||       || ||||  |
Db     973 CCGGTCGGCCTGTTCGTCGAGAC 995


RESULT 7
US-08-403-852D-1
; Sequence 1, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
US-08-403-852D-1

  Query Match             5.6%;  Score 57.4;  DB 2;  Length 5392;
  Best Local Similarity   42.1%;  Pred. No. 0.00053;
  Matches  389;  Conservative    0;  Mismatches  531;  Indels    3;  Gaps    1;

Qy      16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75
           |||  || |  |  |  |||| ||  ||  || |||| | |   |  || || |    |
```

```
Db   3630 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 3689

Qy     76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135

Db   3690 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 3749

Qy    136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195

Db   3750 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 3809

Qy    196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255

Db   3810 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 3869

Qy    256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315

Db   3870 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 3929

Qy    316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372

Db   3930 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 3989

Qy    373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432

Db   3990 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 4049

Qy    433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492

Db   4050 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 4109

Qy    493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552

Db   4110 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCGACATC 4169

Qy    553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612

Db   4170 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 4229

Qy    613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672

Db   4230 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 4289

Qy    673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732

Db   4290 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 4349

Qy    733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792

Db   4350 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 4409

Qy    793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852

Db   4410 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 4469

Qy    853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912

Db   4470 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 4529

Qy    913 GAGGCCGTGGGCAACGGCGAGCC 935

Db   4530 CCGGTCGGCCTGTTCGTCGAGAC 4552


RESULT 8
US-08-510-646B-1
; Sequence 1, Application US/08510646B
; Patent No. 6077699
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/510,646B
;      FILING DATE:  03-AUG-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,852
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-01000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
US-08-510-646B-1

  Query Match             5.6%;  Score 57.4;  DB 3;  Length 5392;
  Best Local Similarity   42.1%;  Pred. No. 0.00053;
  Matches  389;  Conservative    0;  Mismatches  531;  Indels    3;  Gaps    1;

Qy     16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75

Db   3630 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 3689

Qy     76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135

Db   3690 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 3749

Qy    136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195

Db   3750 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 3809

Qy    196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255

Db   3810 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 3869
```

```
Qy     256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
Db    3870 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 3929

Qy     316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372
Db    3930 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 3989

Qy     373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
Db    3990 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 4049

Qy     433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
Db    4050 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 4109

Qy     493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
Db    4110 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCGACATC 4169

Qy     553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
Db    4170 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 4229

Qy     613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
Db    4230 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 4289

Qy     673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
Db    4290 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 4349

Qy     733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
Db    4350 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 4409

Qy     793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
Db    4410 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 4469

Qy     853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
Db    4470 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 4529

Qy     913 GAGGCCGTGGGCAACGGCGAGCC 935
Db    4530 CCGGTCGGCCTGTTCGTCGAGAC 4552
```

```
RESULT 9
US-09-231-818-1
; Sequence 1, Application US/09231818
; Patent No. 6171846
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/231,818
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852
;      FILING DATE:  10-MAY-1995
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5392 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
US-09-231-818-1

  Query Match             5.6%;  Score 57.4;  DB 4;  Length 5392;
  Best Local Similarity   42.1%;  Pred. No. 0.00053;
  Matches  389;  Conservative    0;  Mismatches  531;  Indels    3;  Gaps    1;

Qy      16 GACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCG 75
Db    3630 GACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTCGCGGTCGAGACC 3689

Qy      76 GCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGC 135
Db    3690 CTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAGGCGTACGCGCCC 3749

Qy     136 GGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTC 195
Db    3750 ATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCGTCCGCCAAGGGC 3809

Qy     196 GACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAAC 255
Db    3810 TTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCCCCGGACATCGCC 3869

Qy     256 ATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTG 315
Db    3870 CAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGACGAGCTCGACCAG 3929

Qy     316 GACGCCGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAAGAAGTGCGTGTCGATG 372
Db    3930 CAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAGACCCCCTCGCTG 3989

Qy     373 TCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAG 432
Db    3990 ATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACCGAGGTCCGCAAG 4049

Qy     433 GGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTG 492
Db    4050 GACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACCATCGAGTACCAG 4109
```

```
Qy    493 TGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTC 552
Db   4110 GGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCACGCCGCCGACATC 4169

Qy    553 CGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTC 612
Db   4170 GACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAGCACGTCCTCGCC 4229

Qy    613 ACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCG 672
Db   4230 GCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTGCTGGTCAACCCG 4289

Qy    673 GACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTG 732
Db   4290 ACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACCGGCCGCAAGATC 4349

Qy    733 GCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792
Db   4350 ATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTCTCCGGCAAGGAC 4409

Qy    793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
Db   4410 CCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCCAAGAACGTCGTC 4469

Qy    853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912
Db   4470 GCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATCGGCAAGGCCGAG 4529

Qy    913 GAGGCCGTGGGCAACGGCGAGCC 935
Db   4530 CCGGTCGGCCTGTTCGTCGAGAC 4552


RESULT 10
US-09-023-173-4
; Sequence 4, Application US/09023173
; Patent No. 6066781
;  GENERAL INFORMATION:
;    APPLICANT:  Sutliff, Thomas D.
;    APPLICANT:  Rodriguez, Raymond L.
;    TITLE OF INVENTION:  Production of Mature Proteins
;    TITLE OF INVENTION:  in Plants
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Ave., Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,173
;      FILING DATE:  13-FEB-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/038,168
;      FILING DATE:  13-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petithory, Joanne R
;      REGISTRATION NUMBER:  P42995
;      REFERENCE/DOCKET NUMBER:  0665-0007.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-0880
;      TELEFAX:  650-324-0960
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  codon-optimized RAmy3D signal-prosubtilisin BPN'
US-09-023-173-4

  Query Match             5.6%;  Score 57;  DB 3;  Length 1140;
  Best Local Similarity   46.3%;  Pred. No. 0.00049;
  Matches  232;  Conservative    0;  Mismatches  260;  Indels    9;  Gaps    1;

Qy    447 GTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCC 506
Db    468 GACCAACCCGTTCCAGGACACCAACAGCCATGGCACCCACGTCGCCGGCACCGTCGCCGC 527

Qy    507 CGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCG 566
Db    528 CCTCACCAACAGCATCGGCGTCCTCGGCGTCGCCCCGAGCGCCAGCCTCTACGCCGTCAA 587

Qy    567 CAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTT 626
Db    588 GGTACTCGGCGCCGACGGCAGCGGCCAGTACAGCTGGATCATCAACGGCATCGAGTGGGC 647

Qy    627 CGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGC 686
Db    648 CATCGCCAACAACATGGACGTCATCACCATGAGCCTCGGCGGCCCGAGCGGCAGCGCCGC 707

Qy    687 CGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCG 746
Db    708 CCTCAAGGCCGCCGTCGACAAGGCCGTCGCCAGCGGCGTCGTCGTCGTCGCCGCCGCCGG 767

Qy    747 GAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCAT 806
Db    768 CAACGAGGGCACCAGCGGCAGCAGCAGCACCGTCGGCTACCCGGGCAAGTACCCGAGCGT 827

Qy    807 CGAAGCCACCCTGCGCGACGTCACCAGCCGTGAAC---------TCTACGACGCCGACGA 857
Db    828 CATCGCCGTCGGCGCCGTGGACAGCAGCAACCAGCGCGCGAGCTTCAGCAGCGTCGGCCC 887

Qy    858 GTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGC 917
Db    888 GGAGCTGGACGTCATGGCCCCGGGCGTCAGCATCCAGAGCACCCTCCCGGGCAACAAGTA 947

Qy    918 CGTGGGCAACGGCGAGCCCGG 938
Db    948 CGGCGCCTACAGCGGCACCAG 968


RESULT 11
US-09-173-300-21
; Sequence 21, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT: Hitz, William D.
;  APPLICANT: Kinney, Anthony J.
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/09/173,300
;  CURRENT FILING DATE: 1998-10-15
;  EARLIER APPLICATION NUMBER: 60/063,423
;  EARLIER FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 21
;   LENGTH: 1162
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-173-300-21

  Query Match             5.4%;  Score 55.4;  DB 4;  Length 1162;
  Best Local Similarity   49.9%;  Pred. No. 0.0011;
```

```
  Matches  214;  Conservative    0;  Mismatches  191;  Indels   24;  Gaps    2;

Qy     579 CCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGA 638
           | | ||||| ||||||  | | |           |       |||   | || ||| ||||
Db     267 CGCCACCATGAAGAACGTCAACTACCTCCCCAACGTGCTGTCCATCATGGACGCCGAGGA 326

Qy     639 CCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGG 698
           ||| || ||| | |  || |   | | |||   ||   |  ||| ||||| || ||
Db     327 CCGCGGCGCGTTCGCGTCGGTGTGGGTGGACGGCGAGGGCAACGTCGCCGAGGGGCCCAT 386

Qy     699 CTTCAACGTGTGCATCGTCAAGGACGGCAA------GCTGGCCTCCCCGTCCCGGAACGC 752
             | ||||||    |||||| || || |       | ||  | |  ||| |   ||
Db     387 GGTGAACGTGGCGTTCGTCACGGCCGCCGGCGAGCTGGTGCTCCCGGCGTTCGACAAGAT 446

Qy     753 GTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGC 812
           |  | ||   ||||  |||  | || ||| |||||       | |||         |
Db     447 CCTCGCCGGGTGCACCGCCAAGCGGCTGCTCGCGCTGGCGCCGAGGCTGGTGGAGTCCGG 506

Qy     813 CACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGA------------------ 854
           |  |||    | |||||||| |||  |  ||  || || |||
Db     507 CCTCCTCAAGGCCGTCACCACCCGCCACATCGCCGCCGACGAGGCCAAGCGCTGCTCCGC 566

Qy     855 CGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGA 914
           |||| ||  |   |||  || ||    |  |||   || |||  |  |  ||| | | |
Db     567 CGAGATGGCGTTCGTCGGCAGCGGCCTCCCCGTCCTGCCCATCGTCGAGTGGGACGACCA 626

Qy     915 GGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGGC 974
           |  | | |||  |||| ||   ||       || | ||||  |   ||| | | ||||||
Db     627 GCTCATCGGCGACGGGAAGGTGGGGAAGACGATGATGGCGCTGTCGGATCTGCTCTGGGA 686

Qy     975 GCTGATGGA 983
           |   ||| |
Db     687 GGACATGAA 695


RESULT 12
US-09-025-691-2
; Sequence 2, Application US/09025691
; Patent No. 6069299
;  GENERAL INFORMATION:
;    APPLICANT:  Broadway, Roxanne M.
;    APPLICANT:  Harman, Gary E.
;    TITLE OF INVENTION:  FUNGUS AND INSECT CONTROL WITH
;    TITLE OF INVENTION:  CHITINOLYTIC ENZYMES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/025,691
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/20120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1294 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-025-691-2

  Query Match             5.3%;  Score 53.8;  DB 3;  Length 1294;
  Best Local Similarity   41.9%;  Pred. No. 0.0025;
  Matches  389;  Conservative    0;  Mismatches  537;  Indels    3;  Gaps    1;

Qy      35 TGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGT 94
           ||||   |  | | |||     |   || ||        |||  ||    | |  ||||
Db     196 TGGTACGCACCTACCCCCTTCCGCACCCCGGCCGGCGCCCCTCCACGCCCGGCCTCCACC 255

Qy      95 ACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGG 154
           ||| | |          | | ||  || | ||  || |  ||||| | || || |||
Db     256 GCAGGGGCCGGCTGACCGCCGCCCTCACCGCGGCCGTCCTCGGCGCCTCCGGGCTCGCCC 315

Qy     155 GCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATT 214
           |      ||  || ||| |  | | |  | || |  | |  |||      |
Db     316 TCACCGGCCCCGCGACCGCCGGCGAGGGGGCCCCCGCCGCCCAGGCCGCCCCGGCCGCCG 375

Qy     215 CCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACC 274
                   |         ||   |       | ||  ||||| |  |  |  ||   |
Db     376 TACCGGCCCACGCGGTGACCGGTTACTGGCAGAACTTCAACAACGGCGCGACCGTGCAGA 435

Qy     275 ACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGG 334
           ||| | |    |||  ||||          ||   |  | |   |  ||  ||  || ||
Db     436 CCCTCGCCGACGTGCCGGACGCCTACGACATCATCGCCGTCTCCTTCGCCGACGCCACGG 495

Qy     335 ACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCG 394
           | |   || ||||||||||   |    ||  ||| |   ||       || |  |  ||
Db     496 CCAACGCGGGCGAGATCACCTTCACCCTCGACTCGGTCGGGCTCGGCGGCTACACCGACG 555

Qy     395 TGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGC 454
           | |  |   || |  ||  |     |    ||  | | ||| ||  ||   |   |  |
Db     556 AGCAGTTCCGCGCCGACCTCGCCGCCAAGCAGGCCGACGGCAAGTCGGTGATCATCTCGG 615

Qy     455 TCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGC 514
           ||  |   ||  |  |     |||| ||    ||        ||  || || |  ||
Db     616 TCGGCGGCGAGAAGGGCGCGGTCGCCGTCAACGACAGCGCCTCCGCCCAGCGCTTCGCCG 675

Qy     515 AGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCG 574
           | | |  |  | |      ||| | |  |   ||  | ||| || |  | ||   |
Db     676 ACAGCACCTACGCGCT---GATGGAGGAGTACGGCTTCGACGGCGTCGACATCGACCTGG 732

Qy     575 TCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCA 634
           ||     || |   | |||| |   |  |  ||||||     |       | |||
Db     733 AGAACGGCCTCAACTCCACCTACATGACCGAGGCCCTCACCAAGCTCCACGAGAAGGCCG 792

Qy     635 AGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTC 694
           |||| |    |  | ||||  |   | ||    |   ||||      | ||||
Db     793 GGGACGGCCTGGTCCTCACCATGGCGCCGCAGACCATCGACATGCAGTCGCCCGAGAACG 852

Qy     695 CGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGT 754
           |  ||||||   |    | |||| | | | |   ||| || ||      |  | |    |
Db     853 AGTACTTCAAGACGGCGCTGGTCACGAAAGACTTCCTGACCGCCGTCAACATGCAGTACT 912

Qy     755 TGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCA 814
             ||  | |                 ||  | || | ||  |  ||||| ||  | |
Db     913 ACAACAGCGGCTCGATGCTCGGCTGCGACGGCCAGGTCTACGCGCAGGGCACCGTCGACT 972

Qy     815 CCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCA 874
           |||   ||   || || ||    | ||  |  |||    |||         ||||  ||
Db     973 TCCTCACCGCGCTCGCCTGCATCCAGCTGGAGAACGGTCTCGACGCCTCCCAGGTCGGCA 1032

Qy     875 CCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGC 934
           ||  | |   | | | ||||      |    | ||| | | || |  |  ||  |
Db    1033 TCGGTGTCCCCGCCTCCCCGAAGGCGGCCGGCGGCGGCTACGTCGAGCCCTCCGTGGTCA 1092

Qy     935 CCGGTCCACTGACGGTGGCCATCCGGGAC 963
           ||   | |||         | ||||| |
Db    1093 ACGACGCGCTGGACTGCCTGACCCGGGGC 1121
```

```
RESULT 13
US-08-196-218-31
; Sequence 31, Application US/08196218
; Patent No. 5614619
;  GENERAL INFORMATION:
;    APPLICANT:  Piepersberg, Wolfgang
;    APPLICANT:  Stockmann, Michael
;    APPLICANT:  Taleghani, Kampiz Mansouri
;    APPLICANT:  Distler, Jurgen
;    APPLICANT:  Grabley, Susanne
;    APPLICANT:  Sichel, Petra
;    APPLICANT:  Brau, Barbara
;    TITLE OF INVENTION:  Secondary-Metabolite Biosynthesis Genes
;    TITLE OF INVENTION:  From Actinomycetes, Method of Isolating Them, and Their
;    TITLE OF INVENTION:  Use.
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/196,218
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ogden, Stasia L.
;      REGISTRATION NUMBER:  36,228
;      REFERENCE/DOCKET NUMBER:  02481.1372-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2634 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..401
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  416..1531
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1561..2625
US-08-196-218-31

  Query Match             5.3%;  Score 53.6;  DB 1;  Length 2634;
  Best Local Similarity   43.2%;  Pred. No. 0.0031;
  Matches  430;  Conservative    0;  Mismatches  549;  Indels   17;  Gaps    3;

Qy      18 CCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGC 77
Db    1641 CATCGCCAACACACCCGTGCTGGTGCATGTGCTGAACGCCGTCCGGGAGCTGGGCGTGAC 1700

Qy      78 CGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGG 137
Db    1701 CGAGGTCGGCGTCATCGTCGGCAACCGCGGCCCCGAGATCGAGGCCGTGCTCGGCGACGG 1760

Qy     138 CGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGA 197
Db    1761 TGCCCGGTTCGACGTGCGCATCACCTACATCCCCCAGGACGCACCGCGCGGACTGGCCCA 1820

Qy     198 CACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACAT 257
Db    1821 CACCGTGTCCATCGCCCGCGGCTTCCTCGGCGACGACGACTTCGTGATGTACCTCGGCGA 1880

Qy     258 CTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGA 317
Db    1881 CAACATGCTGCCCGACGGAGTCACCGAGATCGCCGAGGAGTTCACCCGGCAGCGCCCGGC 1940

Qy     318 CGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCA 377
Db    1941 CGCCCAGGTCGTCGTGCACAAGGTCCCCGACCCGCGCTCCTTCGGCGTCGCCGAACTCGG 2000

Qy     378 GCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGA 437
Db    2001 CCCCGACGGGGAGGTGCTGCGCCTGGTGGAGAAGCCGTGGCAGCCGCGCAGCGACATGGC 2060

Qy     438 GAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGC 497
Db    2061 CCTGATCGGGGTCTACTTCTTCACCGCCGCCATCCACCAGGCGGTGGCGGCCATCTCGCC 2120

Qy     498 CTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCG 557
Db    2121 CAGCAGCCGCGGCGAACTGGAGATCACCGACGCCGTCCAGTGGTTGGTCACCTCCGGCGC 2180

Qy     558 CGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGC 617
Db    2181 GGACGTGCGCGCCAGCCTCTACGACGGCTACTGGAAGGACACCGGGAGGG------TCGA 2234

Qy     618 AGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAA 677
Db    2235 GGACGTCCTTGAGTGCAACAGCCACCTCCTGGACGGCCTGACCCCGCGCGTCGACGGACA 2294

Qy     678 CTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTC 737
Db    2295 GGTCGACGCCGACAGCGTGCTCGTCGGCCGGGTCGTGATCGAGGCGGGGGCGCGCATCGT 2354

Qy     738 CCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCA 797
Db    2355 GCGGTCGCGGGTCGAGGGCCCGGCGATCATCGGCGCGGGCACGGTCCTTCAGGACAGCCA 2414

Qy     798 GATGGGCATCGAAGCCACCCT---------GCGCGACGTCACCAGCCGTGAACTCTACGA 848
Db    2415 GGTGGGCCCGCACACCTCCATCGGGCGGGACTGCACGGTGACGGACAGCCGGCTGGAGGG 2474

Qy     849 CGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGA 908
Db    2475 CTCCATCGCCCTGGACGAGGCGTCGGTCACCGGCGTGCGCGGCCTGCGCAACTCGCTGAT 2534

Qy     909 TGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTT 968
Db    2535 CGGGCGCGCCGCGTCCGTCGGCACCACCGGCCCCGGCACG--GGCCATCACTGCCTGGTC 2592

Qy     969 CTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGA 1004
Db    2593 GTCGGAGACCACACCCGAGTGGAGGTCGCGGCATGA 2628


RESULT 14
US-08-681-953-31
; Sequence 31, Application US/08681953
; Patent No. 5710032
;  GENERAL INFORMATION:
;    APPLICANT:  Piepersberg, Wolfgang
;    APPLICANT:  Stockmann, Michael
;    APPLICANT:  Taleghani, Kampiz Mansouri
;    APPLICANT:  Distler, Jurgen
;    APPLICANT:  Grabley, Susanne
;    APPLICANT:  Sichel, Petra
;    APPLICANT:  Brau, Barbara
;    TITLE OF INVENTION:  Secondary-Metabolite Biosynthesis Genes
;    TITLE OF INVENTION:  From Actinomycetes, Method of Isolating Them, and Their
;    TITLE OF INVENTION:  Use.
```

```
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/681,953
;      FILING DATE:  30-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/196,218
;      FILING DATE:  25-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ogden, Stasia L.
;      REGISTRATION NUMBER:  36,228
;      REFERENCE/DOCKET NUMBER:  02481.1372-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2634 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..401
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  416..1531
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1561..2625
US-08-681-953-31

  Query Match             5.3%;  Score 53.6;  DB 1;  Length 2634;
  Best Local Similarity   43.2%;  Pred. No. 0.0031;
  Matches  430;  Conservative    0;  Mismatches  549;  Indels   17;  Gaps    3;

Qy      18 CCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGC 77
           | ||| || | |   | || |||     | ||    |||| |||| |||     |  | |
Db    1641 CATCGCCAACACACCCGTGCTGGTGCATGTGCTGAACGCCGTCCGGGAGCTGGGCGTGAC 1700

Qy      78 CGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGG 137
           ||     |   ||     | || ||  ||  |  || | |     |  | ||| || |||
Db    1701 CGAGGTCGGCGTCATCGTCGGCAACCGCGGCCCCGAGATCGAGGCCGTGCTCGGCGACGG 1760

Qy     138 CGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGA 197
            | |   |   || | ||     |      || | || ||   |  |   |   | | |
Db    1761 TGCCCGGTTCGACGTGCGCATCACCTACATCCCCCAGGACGCACCGCGCGGACTGGCCCA 1820

Qy     198 CACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACAT 257
           |||||  | |         ||   |   |  | || ||     ||   | ||  |  |
Db    1821 CACCGTGTCCATCGCCCGCGGCTTCCTCGGCGACGACGACTTCGTGATGTACCTCGGCGA 1880

Qy     258 CTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGA 317
           |  |  |||   ||||    |   |  | |   ||| | |     ||  || ||||| ||
Db    1881 CAACATGCTGCCCGACGGAGTCACCGAGATCGCCGAGGAGTTCACCCGGCAGCGCCCGGC 1940

Qy     318 CGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCA 377
           ||||  |  |  |  | || |  |  ||||      |       ||||  |
Db    1941 CGCCCAGGTCGTCGTGCACAAGGTCCCCGACCCGCGCTCCTTCGGCGTCGCCGAACTCGG 2000

Qy     378 GCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGA 437
           |    ||    | |  ||   |||   |  |  | | |  |         || |    |
Db    2001 CCCCGACGGGGAGGTGCTGCGCCTGGTGGAGAAGCCGTGGCAGCCGCGCAGCGACATGGC 2060

Qy     438 GAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGC 497
           |  |  |  | |  ||  |     |    |  | | ||  |  ||  | | | | |
Db    2061 CCTGATCGGGGTCTACTTCTTCACCGCCGCCATCCACCAGGCGGTGGCGGCCATCTCGCC 2120

Qy     498 CTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCG 557
           |  |   | || ||| | |    ||  |  | |||       |  |  |     | ||
Db    2121 CAGCAGCCGCGGCGAACTGGAGATCACCGACGCCGTCCAGTGGTTGGTCACCTCCGGCGC 2180

Qy     558 CGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGC 617
           | ||  |||  || | || ||      |  | |||  ||    || | |        ||
Db    2181 GGACGTGCGCGCCAGCCTCTACGACGGCTACTGGAAGGACACCGGGAGGG------TCGA 2234

Qy     618 AGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAA 677
           | |    | ||   |||   ||        | |  |  | || || || || |   |   |
Db    2235 GGACGTCCTTGAGTGCAACAGCCACCTCCTGGACGGCCTGACCCCGCGCGTCGACGGACA 2294

Qy     678 CTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTC 737
             ||  ||||| |   |  | ||  |  |  | | |  || |||  ||   ||    |
Db    2295 GGTCGACGCCGACAGCGTGCTCGTCGGCCGGGTCGTGATCGAGGCGGGGGCGCGCATCGT 2354

Qy     738 CCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCA 797
           | ||| |||  || |   ||||  |||| | |   |      | ||   | || |  |||
Db    2355 GCGGTCGCGGGTCGAGGGCCCGGCGATCATCGGCGCGGGCACGGTCCTTCAGGACAGCCA 2414

Qy     798 GATGGGCATCGAAGCCACCCT---------GCGCGACGTCACCAGCCGTGAACTCTACGA 848
           | |||||    |  || || |           ||   || ||    |  ||    || | |
Db    2415 GGTGGGCCCGCACACCTCCATCGGGCGGGACTGCACGGTGACGGACAGCCGGCTGGAGGG 2474

Qy     849 CGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGA 908
           | ||  ||   ||   | ||   |   | | |||| |  |   |  | ||||||||||
Db    2475 CTCCATCGCCCTGGACGAGGCGTCGGTCACCGGCGTGCGCGGCCTGCGCAACTCGCTGAT 2534

Qy     909 TGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTT 968
           ||    |||| |  |  ||||    |||| ||    |||  |||||||   | | |||
Db    2535 CGGGCGCGCCGCGTCCGTCGGCACCACCGGCCCCGGCACG--GGCCATCACTGCCTGGTC 2592

Qy     969 CTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGA 1004
           | || |   |    ||||  | |  ||| |   ||
Db    2593 GTCGGAGACCACACCCGAGTGGAGGTCGCGGCATGA 2628


RESULT 15
US-09-003-287-7
; Sequence 7, Application US/09003287
; Patent No. 6096947
; GENERAL INFORMATION:
;  APPLICANT: Jayne, Susan
;  APPLICANT: Barbour, Eric
;  APPLICANT: Meyer, Terry
;  TITLE OF INVENTION: METHODS FOR IMPROVING TRANSFORMATION EFFICIENCY
;  FILE REFERENCE: moPAT_moCAH
;  CURRENT APPLICATION NUMBER: US/09/003,287
;  CURRENT FILING DATE: 1998-01-06
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 735
;   TYPE: DNA
;   ORGANISM: Myrothecium verrucaria
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (01)..(732)
US-09-003-287-7

  Query Match             5.2%;  Score 53.2;  DB 3;  Length 735;
  Best Local Similarity   43.7%;  Pred. No. 0.003;
  Matches  235;  Conservative    0;  Mismatches  303;  Indels    0;  Gaps    0;
```

```
Qy    315 GGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTC 374
          || |  || | ||  |    | | | || | |||  ||| |    | | ||||  |
Db     57 GGCCATCGTGGACTCCCTCGGCAAGCTCGGCGACGTCAGCTCCTACTCCGTGGAGGACAT 116

Qy    375 GCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGG 434
            | |  |    |   |  |    || |||   ||| ||  | ||  |||  |
Db    117 CGCCTTCCCAGCGGCCGACAAGCTCGTGGCCGAGGCCCAGGCGTTCGTCAAGGCCCGCCT 176

Qy    435 CGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTG 494
          |      |||     | ||| | | ||    |  |||| |  ||  | || |  |  |    |
Db    177 CTCCCCGGAGACCTACAACCACTCCATGCGCGTGTTCTACTGGGGCACCGTGATCGCCAG 236

Qy    495 GGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCG 554
          |    |||  || |   ||   |    |  | || |  | | |  | | | |     ||
Db    237 GAGGCTCCTCCCGGAGCAGGCCAAGGACCTCTCCCCGTCCACCTGGGCCCTCACCTGCCT 296

Qy    555 CCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCAC 614
          || |  || |     ||||| |||  |   | |||     |  ||     |         ||
Db    297 CCTCCACGACGTGGGCACCGCCGAGGCCTACTTCACCTCCACCCGCATGTCCTTCGACAT 356

Qy    615 CGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGA 674
          |   | | |  || | || | |||         |     ||    |    ||||  ||
Db    357 CTACGGCGGCATCAAGGCGATGGAGGTCCTCAAGGTGCTCGGCTCCTCCACCGACCAGGC 416

Qy    675 CAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGC 734
            |   |||||||| |        |  | ||| |  | ||| |   ||||||||  |  |
Db    417 GGAGGCGGTGGCCGAGGCCATCATCCGCCACGAGGACGTCGGCGTCGACGGCAACATCAC 476

Qy    735 CTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGA 794
          || |-|    || |  |     ||  | || |  |    | |  |||  ||     | ||
Db    477 CTTCCTCGGCCAGCTCATCCAGCTCGCCACCCTCTACGACAACGTGGGCGCCTACGACGG 536

Qy    795 CCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852
          |     | | |||   ||    ||  |||| |||| ||       || ||  ||||
Db    537 CATCGACGACTTCGGCTCCTGGGTGGACGACACCACCCGCAACTCCATCAACACCGCC 594
```

Search completed: April 24, 2003, 20:08:52
Job time : 137 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 20:08:57 ; Search time 136 Seconds
(without alignments)
8160.966 Million cell updates/sec

Title: US-10-067-291-2
Perfect score: 1020
Sequence: 1 atgactgctctttcagacct.....tcgaaacgatcgaatactga 1020

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1020	100.0	1020	12	US-10-067-291-2	Sequence 2, Appli
2	325	31.9	993	9	US-10-112-706-6	Sequence 6, Appli
3	325	31.9	993	9	US-10-112-706-7	Sequence 7, Appli
4	325	31.9	1574	9	US-10-112-706-21	Sequence 21, Appl
5	324.6	31.8	978	9	US-10-112-706-5	Sequence 5, Appli
6	164.8	16.2	465	9	US-10-112-706-26	Sequence 26, Appl
7	66.2	6.5	15872	9	US-09-860-846-1	Sequence 1, Appli
8	66.2	6.5	15872	9	US-09-988-384B-1	Sequence 1, Appli
9	66.2	6.5	15872	10	US-09-861-289-1	Sequence 1, Appli
10	56.2	5.5	6798	9	US-09-918-740-57	Sequence 57, Appl
11	56.2	5.5	8077	9	US-09-918-740-63	Sequence 63, Appl
12	55.4	5.4	1162	12	US-10-027-450-21	Sequence 21, Appl
13	53.8	5.3	1294	10	US-09-748-033-2	Sequence 2, Appli
14	52.4	5.1	810	10	US-09-974-300-492	Sequence 492, App
15	51	5.0	1266	9	US-10-145-415-98	Sequence 98, Appl
16	50.4	4.9	1086	12	US-10-027-450-14	Sequence 14, Appl
17	50.2	4.9	1835	10	US-09-779-429-2	Sequence 2, Appli
18	49.2	4.8	905	10	US-09-748-033-5	Sequence 5, Appli
19	48.6	4.8	3126	10	US-09-815-242-4020	Sequence 4020, Ap
20	48.4	4.7	88421	9	US-09-976-059-1	Sequence 1, Appli
21	48.2	4.7	1473	10	US-09-070-844-25	Sequence 25, Appl
22	48.2	4.7	1506	10	US-09-070-844-23	Sequence 23, Appl
23	48.2	4.7	1969	10	US-09-070-844-7	Sequence 7, Appli
24	48.2	4.7	2096	10	US-09-070-844-19	Sequence 19, Appl
25	48.2	4.7	2099	10	US-09-070-844-3	Sequence 3, Appli
26	48.2	4.7	2137	10	US-09-070-844-18	Sequence 18, Appl
27	48.2	4.7	2140	10	US-09-070-844-1	Sequence 1, Appli
28	47	4.6	390	10	US-09-790-399-7	Sequence 7, Appli
29	46.8	4.6	822	9	US-10-101-464A-38	Sequence 38, Appl
30	46.8	4.6	1248	9	US-09-860-846-7	Sequence 7, Appli
31	46.8	4.6	1248	9	US-09-988-384B-7	Sequence 7, Appli
32	46.8	4.6	1248	10	US-09-861-289-7	Sequence 7, Appli
33	46.8	4.6	2138	9	US-10-101-464A-462	Sequence 462, App
c 34	46.8	4.6	12441	9	US-09-988-384B-3	Sequence 3, Appli
c 35	46.8	4.6	13613	9	US-09-860-846-3	Sequence 3, Appli
c 36	46.8	4.6	13613	10	US-09-861-289-3	Sequence 3, Appli
37	46.4	4.5	2451	10	US-09-815-242-7752	Sequence 7752, Ap
38	46.2	4.5	1503	10	US-09-815-242-7680	Sequence 7680, Ap
39	45.6	4.5	1185	10	US-09-887-576-784	Sequence 784, App
40	45.6	4.5	1609	9	US-09-931-457A-50	Sequence 50, Appl
41	45.6	4.5	8911	12	US-10-044-090-17	Sequence 17, Appl
42	45.4	4.5	1185	10	US-09-815-242-7687	Sequence 7687, Ap
43	45.4	4.5	4041	9	US-09-860-846-36	Sequence 36, Appl
44	45.4	4.5	4041	9	US-09-988-384B-36	Sequence 36, Appl
45	45.4	4.5	4041	10	US-09-861-289-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-10-067-291-2
; Sequence 2, Application US/10067291
; Patent No. US20020127664A1
; GENERAL INFORMATION:
; APPLICANT: Takashima, Yoshiki
; APPLICANT: Mitsuda, Satoshi
; TITLE OF INVENTION: Protein capable of catalyzing transamination
; TITLE OF INVENTION: stereoselectively, gene encoding said protein
; TITLE OF INVENTION: and use thereof
; FILE REFERENCE: Q58251
; CURRENT APPLICATION NUMBER: US/10/067,291
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 09/527,522
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: JP 11/075511
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 11/088634
; PRIOR FILING DATE: 1999-03-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Mycobacterium aurum
; FEATURE:
; OTHER INFORMATION: SC-S423
; NAME/KEY: CDS
; LOCATION: (1)..(1017)
US-10-067-291-2

Query Match 100.0%; Score 1020; DB 12; Length 1020;
Best Local Similarity 100.0%; Pred. No. 5e-253;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGACTGCTCTTTCAGACCTCGGCACCTCCAACCTGGTGGCCGTCGAGCCCGGCGCCATC 60

Qy      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACCTCC 120
```

```
Qy    121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGGCCGAAGAAGCG 180

Qy    181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACACCGTCGCGCAT 240

Qy    241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCG 300

Qy    301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAG 360

Qy    361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATAC 420

Qy    421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCC 480

Qy    481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTG 540

Qy    541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAG 600

Qy    601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATC 660

Qy    661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAG 720

Qy    721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTG 780

Qy    781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAA 840

Qy    841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAAC 900

Qy    901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGG 960

Qy    961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGATCGAATACTGA 1020
```

RESULT 2
US-10-112-706-6
; Sequence 6, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786A1iyuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: OGURA, Masahiro
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/403,493
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-121732
; PRIOR FILING DATE: 1997-04-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 993
; TYPE: DNA
; ORGANISM: Arthrobacter sp.
US-10-112-706-6

Query Match 31.9%; Score 325; DB 9; Length 993;
Best Local Similarity 58.2%; Pred. No. 1.7e-74;
Matches 571; Conservative 0; Mismatches 410; Indels 0; Gaps 0;

```
Qy     40 GCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGC 99
          ||||      |||    | ||  |  | ||  | | | |  |       |||   || |||
Db     13 GCCGATACCTCCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGC 72

Qy    100 GACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAA 159
          ||||||||||| ||  |  | | |||| |||| || || |  || ||||||||||| | |
Db     73 GACTACGAACTCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCA 132

Qy    160 TACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGAT 219
          | | ||||| |    || |||  |||||| ||||||||    || | |   || || ||
Db    133 TTCGTGCCGCCGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGAC 192

Qy    220 CTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTG 279
          | |||||||| |||   || || ||| |||| ||    ||||| |||| |||||||| |
Db    193 GTCACCTACACGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATC 252

Qy    280 GACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAA 339
          || ||  | ||  |  |||     | ||||| |  | |   |  ||  ||||||||
Db    253 GAACGCCTCTTCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAA 312

Qy    340 CTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAAT 399
          ||   |||||  |    |  |||  ||| | |  | |||| ||  | |||||
Db    313 GTGAAGGAGATTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCC 372

Qy    400 CTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACC 459
          || |  | ||||| || ||| |  |     |||||||  ||| |  | |||| |  |
Db    373 GTGTCGATTACCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGC 432

Qy    460 CATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATC 519
          |  |||||||||| || ||| |||| || | ||||  | | |||   | |        |
Db    433 CCGCAGGTGTACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGC 492

Qy    520 TTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGAC 579
          ||||    ||| || ||  | |     || |||||  ||   |||| | |  |||||
Db    493 GACGGCGTGCACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGAC 552

Qy    580 CCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGAC 639
          ||     |||||||| |||||||| ||||| | |   ||    |   ||  |  | |||
Db    553 CCTCAGGTCAAGAACTTCCAGTGGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGAC 612

Qy    640 CGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGC 699
          || ||        |  | | |||||||||   ||          | ||||| ||  ||||
Db    613 CGCGGGTTCGAGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGG 672

Qy    700 TTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCG 759
          ||||||||   | |  ||||||||||||     ||  |  ||||  ||    ||| | ||
Db    673 TTCAACGTCGTCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCC 732

Qy    760 GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTG 819
          ||||| || || ||||| ||| ||||  | |||||   | | ||   ||| || |  ||
Db    733 GGCATTACGCGGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTC 792
```

```
Qy    820 CGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCG 879
Db    793 GCCGACATCACGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCG 852

Qy    880 GGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGT 939
Db    853 GGCGGAGTGTGGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGC 912

Qy    940 CCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTG 999
Db    913 CCCATCACCCAGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCGTCGTTG 972

Qy   1000 ATCGAAACGATCGAATACTGA 1020
Db    973 CTTACGCCTGTGCAGTACTGA 993
```

RESULT 3
US-10-112-706-7
; Sequence 7, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786Aliyuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: OGURA, Masahiro
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/403,493
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-121732
; PRIOR FILING DATE: 1997-04-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 993
; TYPE: DNA
; ORGANISM: Arthrobacter sp.
US-10-112-706-7

Query Match 31.9%; Score 325; DB 9; Length 993;
Best Local Similarity 58.2%; Pred. No. 1.7e-74;
Matches 571; Conservative 0; Mismatches 410; Indels 0; Gaps 0;

```
Qy     40 GCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGC 99
Db     13 GCCGATACCTCCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGC 72

Qy    100 GACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAA 159
Db     73 GACTACGAACTCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCA 132

Qy    160 TACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGAT 219
Db    133 TTCGTGCCGCCGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGAC 192

Qy    220 CTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTG 279
Db    193 GTCACCTACACGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATC 252

Qy    280 GACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAA 339
Db    253 GAACGCCTCTTCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAA 312

Qy    340 CTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAAT 399
Db    313 GTGAAGGAGATTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCC 372

Qy    400 CTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACC 459
Db    373 GTGTCGATTACCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGC 432

Qy    460 CATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATC 519
Db    433 CCGCAGGTGTACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGC 492

Qy    520 TTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGAC 579
Db    493 GACGGCGTGCACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGAC 552

Qy    580 CCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGAC 639
Db    553 CCTCAGGTCAAGAACTTCCAGTGGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGAC 612

Qy    640 CGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGC 699
Db    613 CGCGGGTTCGAGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGG 672

Qy    700 TTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCG 759
Db    673 TTCAACGTCGTCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCC 732

Qy    760 GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTG 819
Db    733 GGCATTACGCGGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTC 792

Qy    820 CGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCG 879
Db    793 GCCGACATCACGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCG 852

Qy    880 GGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGT 939
Db    853 GGCGGAGTGTGGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGC 912

Qy    940 CCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTG 999
Db    913 CCCATCACCCAGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCGTCGTTG 972

Qy   1000 ATCGAAACGATCGAATACTGA 1020
Db    973 CTTACGCCTGTGCAGTACTGA 993
```

RESULT 4
US-10-112-706-21
; Sequence 21, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786Aliyuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: OGURA, Masahiro
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/403,493
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-121732
; PRIOR FILING DATE: 1997-04-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1574
; TYPE: DNA
; ORGANISM: Arthrobacter sp.
US-10-112-706-21

Query Match 31.9%; Score 325; DB 9; Length 1574;
Best Local Similarity 58.2%; Pred. No. 1.7e-74;
Matches 571; Conservative 0; Mismatches 410; Indels 0; Gaps 0;

```
Qy      40 GCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGC 99
           ||||      |||    | ||  |  | ||  | | | |  |       |||   || |||
Db     398 GCCGATACCTCCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGC 457

Qy     100 GACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAA 159
           |||||||||||| ||   | | |||| |||| || || |  || |||||||||| | |
Db     458 GACTACGAACTCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCA 517

Qy     160 TACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGAT 219
           | | ||||| |    || ||| |||||| ||||||||    || | |   || || ||
Db     518 TTCGTGCCGCCGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGAC 577

Qy     220 CTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTG 279
           | ||||||||| |||   || || ||| |||| ||    ||||| |||| ||||||| |
Db     578 GTCACCTACACGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATC 637

Qy     280 GACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAA 339
           || ||  |  ||  |   |||     | ||||| |   | |   |  ||   ||||||||
Db     638 GAACGCCTCTTCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAA 697

Qy     340 CTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAAT 399
           ||   |||||  |      |   |||  ||| | |  |  |||| ||  | ||||||
Db     698 GTGAAGGAGATTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCC 757

Qy     400 CTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACC 459
           || |  | ||||| || ||| |   |     |||||||   ||| |  | |||| |  |
Db     758 GTGTCGATTACCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGC 817

Qy     460 CATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATC 519
           |  ||||||||||| || ||| |||| || | |||  | | |||   | |         |
Db     818 CCGCAGGTGTACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGC 877

Qy     520 TTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGAC 579
             ||||     ||| || ||  | |     || |||||  ||   |||| | |  |||||
Db     878 GACGGCGTGCACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGAC 937

Qy     580 CCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGAC 639
           ||     |||||||| |||||||||   ||        | ||||| ||  ||||
Db     938 CCTCAGGTCAAGAACTTCCAGTGGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGAC 997

Qy     640 CGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGC 699
           || ||        |  |  |||||||||   ||        | ||||| ||  ||||
Db     998 CGCGGGTTCGAGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGG 1057

Qy     700 TTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCG 759
           ||||||||   | |  |||||||||||     ||  |  ||||  ||    ||| | ||
Db    1058 TTCAACGTCGTCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCC 1117

Qy     760 GGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTG 819
           ||||| || || ||||| ||| |||| | |||||   | | ||   ||| || |  ||
Db    1118 GGCATTACGCGGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTC 1177

Qy     820 CGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCG 879
             |||| ||||    |  ||||||   ||||||||||||  || | |    ||| || |||
Db    1178 GCCGACATCACGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCG 1237

Qy     880 GGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGT 939
           ||||| ||    ||  ||  |   |||| ||| |  || |     |||| |  |||||
Db    1238 GGCGGAGTGTGGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGC 1297

Qy     940 CCACTGACGGTGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTG 999
           ||  | ||   | | |||      || | ||||| |||||  | ||||         || ||
Db    1298 CCCATCACCCAGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCGTCGTTG 1357

Qy    1000 ATCGAAACGATCGAATACTGA 1020
           |    |  |  ||||||
Db    1358 CTTACGCCTGTGCAGTACTGA 1378
```

RESULT 5
US-10-112-706-5
; Sequence 5, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786A1iyuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: OGURA, Masahiro
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/403,493
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-121732
; PRIOR FILING DATE: 1997-04-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 978
; TYPE: DNA
; ORGANISM: Arthrobacter sp.
US-10-112-706-5

Query Match 31.8%; Score 324.6; DB 9; Length 978;
Best Local Similarity 58.4%; Pred. No. 2.1e-74;
Matches 567; Conservative 0; Mismatches 404; Indels 0; Gaps 0;

```
Qy      50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109
           |||    | ||  |  | ||  | | | |  |       |||   || ||||||||||||
Db       8 CCGAGATCGTCTACACGCACGACACCGGCCTCGACTACATCACTTATAGCGACTACGAAC 67

Qy     110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169
           | ||  |  | | |||| |||| || || |  || ||||||||||| | || | |||||
Db      68 TCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGGATCGAGGGTGCATTCGTGCCGC 127

Qy     170 CCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCTACA 229
           |    || |||  |||||| ||||||||    || | |   || || ||  | ||||||||
Db     128 CGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTACCTCCACTCGGACGTCACCTACA 187

Qy     230 CCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACCACCTGGACCGGTTGC 289
           | |||   || || ||| |||| ||    ||||| |||| |||||||| | || ||  |
Db     188 CGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTCGACGACCACATCGAACGCCTCT 247

Qy     290 TCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGGCCGAGA 349
           ||  |  |||    | ||||| |   | |   |  ||  |||||||| ||   ||||
Db     248 TCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTCACACAGGACGAAGTGAAGGAGA 307

Qy     350 TCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGACCGTCA 409
           | |      |  ||| ||| | |  | |||| ||  | ||||||    || |  | |
Db     308 TTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAGGCCTTCGTGTCCGTGTCGATTA 367

Qy     410 CCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATCAGGTGT 469
           |||| || ||| |    |      |||||||   ||| |  | |||| |  || |||||||
Db     368 CCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATCACGAAGCACCGCCCGCAGGTGT 427

Qy     470 ACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCGGCACCA 529
           |||| || ||| |||| || | ||||  | | |||   | |       |  ||||
Db     428 ACATGTATGCCGTCCCATATCAGTGGATCGTGCCGTTTGACCGAATTCGCGACGGCGTGC 487

Qy     530 CCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGACCATCA 589
           ||| || ||  | |      || |||||  ||   |||| | |  |||||||     |||
Db     488 ACGCCATGGTCGCACAGAGCGTGCGCCGAACCCCGCGCAGCTCGATCGACCCTCAGGTCA 547

Qy     590 AGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTGGTGCGC 649
           |||||| |||||||||| |||||  | |    ||    |   ||  |  | |||||  ||
Db     548 AGAACTTCCAGTGGGGGGATCTGATCCGTGCGGTTCAAGAGACGCACGACCGCGGGTTCG 607

Qy     650 GCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCAACGTGT 709
```

```
Db    608 AGGCTCCCCTTCTGCTCGACGGCGATGGACTGCTTGCCGAGGGCTCGGGGTTCAACGTCG 667

Qy    710 GCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCATCACCC 769

Db    668 TCGTGATCAAGGACGGCGTCGTGCGCAGCCCGGGTCGAGCGGCGCTCCCCGGCATTACGC 727

Qy    770 GTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCA 829

Db    728 GGAAGACCGTGCTCGAGATCGCCGAATCGCTCGGACACGAGGCGATTCTCGCCGACATCA 787

Qy    830 CCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCA 889

Db    788 CGCTCGCTGAACTGCTCGACGCCGACGAAGTGCTCGGCTGCACGACTGCGGGCGGAGTGT 847

Qy    890 CACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGG 949

Db    848 GGCCATTCGTCAGCGTGGACGGCAACCCCATCTCGGACGGGGTTCCCGGCCCCATCACCC 907

Qy    950 TGGCCATCCGGGACCGGTTCTGGGCGCTGATGGACGAGCCGGGCCCGCTGATCGAAACGA 1009

Db    908 AGTCGATCATCCGTCGTTACTGGGAGCTGAATGTCGAGAGCTCGTCGTTGCTTACGCCTG 967

Qy   1010 TCGAATACTGA 1020

Db    968 TGCAGTACTGA 978
```

RESULT 6
US-10-112-706-26
; Sequence 26, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786A1iyuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: OGURA, Masahiro
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/403,493
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-121732
; PRIOR FILING DATE: 1997-04-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Arthrobacter sp.
US-10-112-706-26

Query Match 16.2%; Score 164.8; DB 9; Length 465;
Best Local Similarity 62.7%; Pred. No. 2.3e-33;
Matches 256; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

```
Qy     88 ATCCAGTACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGG 147

Db     43 ATCACTTATAGCGACTACGAACTCGATCCTGCTAACCCGCTCGCGGGAGGTGCGGCATGG 102

Qy    148 ATCGAGGGCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTC 207

Db    103 ATCGAGGGTGCATTCGTGCCGCCGTCGGAGGCGCGGATCTCGATCTTCGATCAGGGTTAC 162

Qy    208 GGTCATTCCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTC 267

Db    163 CTCCACTCGGACGTCACCTACACGGTCTTCCACGTCTGGAACGGAAATGCATTCCGCCTC 222

Qy    268 GGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTAC 327

Db    223 GACGACCACATCGAACGCCTCTTCTCCAACGCGGAGTCGATGCGCATCATCCCTCCGCTC 282

Qy    328 AGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAA 387

Db    283 ACACAGGACGAAGTGAAGGAGATTGCGCTCGAACTCGTCGCGAAGACCGAATTGCGTGAG 342

Qy    388 TCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTG 447

Db    343 GCCTTCGTGTCCGTGTCGATTACCCGCGGTTACAGCTCGACTCCGGGCGAGCGCGACATC 402

Qy    448 TCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGG 495

Db    403 ACGAAGCACCGCCCGCAGGTGTACATGTATGCCGTCCCATACCAGTGG 450
```

RESULT 7
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match 6.5%; Score 66.2; DB 9; Length 15872;
Best Local Similarity 44.4%; Pred. No. 5.6e-08;
Matches 358; Conservative 0; Mismatches 443; Indels 6; Gaps 2;

```
Qy     50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109

Db    765 CCCGCGCCCTCGCCGACGGCGACCGTGTCCACGGCGTCATCCGCGCCAGCGCCGTCAACA 824

Qy    110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169

Db    825 ACGACGGAGCCACCCCGGGTCTCACCGTGCCCAGCAGGGCCGCCCAGGAGAAGGTGCTGC 884

Qy    170 CCGAAG---AAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCT 226

Db    885 GCGAGGCGTACCGGAAGGCGGCCCTGGACCCGTCCGCCGTCCAGTACGTCGAACTCCACG 944

Qy    227 ACACCGTCGCGCATGTATGGCACGGCA---ACATCTTCCGGCTCGGCGACCACCTGGACC 283

Db    945 GCACCGGAACCCCCGTCGGCGACCCCATCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCT 1004

Qy    284 GGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGG 343

Db   1005 CGGCGCGCCCCGCGGACGAACCCCTGCTCGTCGGCTCGGCCAAGACGAACGTCGGGCACC 1064

Qy    344 CCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGA 403

Db   1065 TCGAAGGCGCCGCCGGCATCGTCGGCCTCATCAAGACGCTCCTCGCGCTCGGCCGGCGCC 1124

Qy    404 CCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATC 463

Db   1125 GGATCCCGGCGAGCCTCAACTTCCGTACGCCCCACCCGGACATCCCGCTCGACACCCTCG 1184

Qy    464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523
```

```
Db   1185 GGCTCGACGTGCCCGACGGCCTGCGGGAGTGGCCGCACCCGGACCGCGAACTCCTCGCCG 1244

Qy    524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583

Db   1245 GCGTCAGCTCGTTCGGCATGGGCGGCACCAACGCCCACGTCGTCCTCAGCGAAGGCCCCG 1304

Qy    584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643

Db   1305 CCCAGGGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCCGTCGACAGCGGGGCCGCAC 1364

Qy    644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703

Db   1365 TGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGCGCCCAGGCCCGGCGCCTGC 1424

Qy    704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763

Db   1425 ACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCACTCGCCCGGTCGCTGGTCA 1484

Qy    764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823

Db   1485 CCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCC 1544

Qy    824 ACGTCACCAGCCGTGAACTCTACGACG 850

Db   1545 TCGACGGCCTCGGCGCCCTCGCCGCCG 1571


RESULT 8
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

  Query Match             6.5%;  Score 66.2;  DB 9;  Length 15872;
  Best Local Similarity   44.4%;  Pred. No. 5.6e-08;
  Matches  358;  Conservative    0;  Mismatches  443;  Indels    6;  Gaps    2;

Qy      50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109

Db     765 CCCGCGCCCTCGCCGACGGCGACCGTGTCCACGGCGTCATCCGCGCCAGCGCCGTCAACA 824

Qy     110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169

Db     825 ACGACGGAGCCACCCCGGGTCTCACCGTGCCCAGCAGGGCCGCCCAGGAGAAGGTGCTGC 884

Qy     170 CCGAAG---AAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCT 226

Db     885 GCGAGGCGTACCGGAAGGCGGCCCTGGACCCGTCCGCCGTCCAGTACGTCGAACTCCACG 944

Qy     227 ACACCGTCGCGCATGTATGGCACGGCA---ACATCTTCCGGCTCGGCGACCACCTGGACC 283

Db     945 GCACCGGAACCCCCGTCGGCGACCCCATCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCT 1004

Qy     284 GGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGG 343
```

```
Db   1005 CGGCGCGCCCCGCGGACGAACCCCTGCTCGTCGGCTCGGCCAAGACGAACGTCGGGCACC 1064

Qy    344 CCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGA 403

Db   1065 TCGAAGGCGCCGCCGGCATCGTCGGCCTCATCAAGACGCTCCTCGCGCTCGGCCGGCGCC 1124

Qy    404 CCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATC 463

Db   1125 GGATCCCGGCGAGCCTCAACTTCCGTACGCCCCACCCGGACATCCCGCTCGACACCCTCG 1184

Qy    464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523

Db   1185 GGCTCGACGTGCCCGACGGCCTGCGGGAGTGGCCGCACCCGGACCGCGAACTCCTCGCCG 1244

Qy    524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583

Db   1245 GCGTCAGCTCGTTCGGCATGGGCGGCACCAACGCCCACGTCGTCCTCAGCGAAGGCCCCG 1304

Qy    584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643

Db   1305 CCCAGGGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCCGTCGACAGCGGGGCCGCAC 1364

Qy    644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703

Db   1365 TGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGCGCCCAGGCCCGGCGCCTGC 1424

Qy    704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763

Db   1425 ACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCACTCGCCCGGTCGCTGGTCA 1484

Qy    764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823

Db   1485 CCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCC 1544

Qy    824 ACGTCACCAGCCGTGAACTCTACGACG 850

Db   1545 TCGACGGCCTCGGCGCCCTCGCCGCCG 1571


RESULT 9
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-1

  Query Match             6.5%;  Score 66.2;  DB 10;  Length 15872;
  Best Local Similarity   44.4%;  Pred. No. 5.6e-08;
  Matches  358;  Conservative    0;  Mismatches  443;  Indels    6;  Gaps    2;

Qy      50 CCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGTACAGCGACTACGAAC 109

Db     765 CCCGCGCCCTCGCCGACGGCGACCGTGTCCACGGCGTCATCCGCGCCAGCGCCGTCAACA 824

Qy     110 TGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGGGCGAATACCTGCCGG 169

Db     825 ACGACGGAGCCACCCCGGGTCTCACCGTGCCCAGCAGGGCCGCCCAGGAGAAGGTGCTGC 884
```

```
Qy    170 CCGAAG---AAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATTCCGATCTGACCT 226
          |||  |   |   ||||    ||  |   |    |||   ||     |   |||  ||   |
Db    885 GCGAGGCGTACCGGAAGGCGGCCCTGGACCCGTCCGCCGTCCAGTACGTCGAACTCCACG 944

Qy    227 ACACCGTCGCGCATGTATGGCACGGCA---ACATCTTCCGGCTCGGCGACCACCTGGACC 283
          |||||   | |  ||  |  ||  ||    |   |  |||||||| |  |||  | |
Db    945 GCACCGGAACCCCCGTCGGCGACCCCATCGAGGCCGCCGCGCTCGGCGCCGTCCTCGGCT 1004

Qy    284 GGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGGACGAACTGG 343
          |  || |  || ||       |||| | | | | |   | |  || | ||  |
Db   1005 CGGCGCGCCCCGCGGACGAACCCCTGCTCGTCGGCTCGGCCAAGACGAACGTCGGGCACC 1064

Qy    344 CCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCGTGAATCTGA 403
          |||   | ||         |||      |    ||  || |    || |||      |
Db   1065 TCGAAGGCGCCGCCGGCATCGTCGGCCTCATCAAGACGCTCCTCGCGCTCGGCCGGCGCC 1124

Qy    404 CCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGCTCACCCATC 463
          || |   | |   |       || | |  | |   |||| |  |       |||||
Db   1125 GGATCCCGGCGAGCCTCAACTTCCGTACGCCCCACCCGGACATCCCGCTCGACACCCTCG 1184

Qy    464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523
          | |  || |   || |  || |     |||| |   ||||  |  ||| |   ||  ||
Db   1185 GGCTCGACGTGCCCGACGGCCTGCGGGAGTGGCCGCACCCGGACCGCGAACTCCTCGCCG 1244

Qy    524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583
          ||  || | || |||    | ||     |  |  |  || |  |  || ||  | |||
Db   1245 GCGTCAGCTCGTTCGGCATGGGCGGCACCAACGCCCACGTCGTCCTCAGCGAAGGCCCCG 1304

Qy    584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643
          ||        | | |||   ||          |  ||   ||||  |  || |
Db   1305 CCCAGGGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCCGTCGACAGCGGGGCCGCAC 1364

Qy    644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703
          |  |  ||  | |  | ||  |   ||   |      ||| | |  | | || |
Db   1365 TGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGCGCCCAGGCCCGGCGCCTGC 1424

Qy    704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763
          |||  |  | |||    ||||    |||| ||  ||  |  | |    |  || || || ||
Db   1425 ACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCACTCGCCCGGTCGCTGGTCA 1484

Qy    764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823
          ||||||||  |     |  | | | | | |   | | |   ||   | |||  | |
Db   1485 CCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCCCCGGACCGCGCCCGCCTCC 1544

Qy    824 ACGTCACCAGCCGTGAACTCTACGACG 850
          || |  |  | | |  |||  || ||
Db   1545 TCGACGGCCTCGGCGCCCTCGCCGCCG 1571
```

RESULT 10
US-09-918-740-57
; Sequence 57, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways t
; TITLE OF INVENTION: create novel traits in transgenic organisms
; FILE REFERENCE: KAS-103XC1
; CURRENT APPLICATION NUMBER: US/09/918,740
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
; LENGTH: 6798
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Streptomyces sp CL190 gene cluster containing mevalonate pathway
; OTHER INFORMATION: and IPP isomerase orfs

US-09-918-740-57

Query Match 5.5%; Score 56.2; DB 9; Length 6798;
Best Local Similarity 45.9%; Pred. No. 2e-05;
Matches 228; Conservative 0; Mismatches 268; Indels 1; Gaps 1;

```
Qy    464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523
          | | | || ||  | ||  | |  | | | | ||    |   |  ||| | |    | |
Db   1697 AAGCGGACCTCGGCTCCTACGCCGAGCCGGTGCCCGCGGCCGACCTCGACCCGGCGCTGG 1756

Qy    524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583
           || | ||| | ||    |  |||     | |  | || ||||| |  || |  || |
Db   1757 TCATCGCCGTGGTCAACGCCGGCCCCAAGCCCGTCTCCAGCCGCGAGGCCATGCGCCGCA 1816

Qy    584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643
          || || | | ||  | |  |       |      || | |       | | | |       |
Db   1817 CCGTCGACACCTCGCCGCTGTACCGGCCGTGGGCCGACTCCAGTAAGGACGACCTGGACG 1876

Qy    644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703
          ||||  | ||| | |||| || |      | |   ||||| |||      |||    ||
Db   1877 AGATGCGCTCGGCGCTGCTGCGCGGCGACCTCGAGGCCGTGGGCGAGATCGCGGAGCGCA 1936

Qy    704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763
          ||| |  |  | |  | | |  || |||||| |||   ||  |     ||  | |
Db   1937 ACGCGCTCGGCATGCACGCCACCATGCTGGCCGCCCGCCCCGCGGTGCGGTACCTGTCGC 1996

Qy    764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823
            ||      || ||| ||||   |    |||| |  |||  || |   ||||  |
Db   1997 CGGCCACGGTCACCGTGCTCGACAGCGTGCTCCAGCTCCGCAAGGACGGTGTCCTGGCCT 2056

Qy    824 ACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTG-ATGGCGGTCACCACCGCGGGC 882
          |||  ||||     |        | |||   | | | ||    |  |     ||  || ||
Db   2057 ACGCGACCATGGACGCCGGTCCCAACGTGAAGGTGCTGTGCCGGCGGGCGGACGCCGAGC 2116

Qy    883 GGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCA 942
          |||    |  ||   || ||| |    |   ||| ||  |   |     ||| || |
Db   2117 GGGTGGCCGACGTCGTACGCGCCGCCGCGTCCGGCGGTCAGGTCCTCGTCGCCGGGCCGG 2176

Qy    943 CTGACGGTGGCCATCCG 959
            |||||||  ||  | |
Db   2177 GAGACGGTGCCCGCCTG 2193
```

RESULT 11
US-09-918-740-63
; Sequence 63, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; APPLICANT: Kuehnle, Adelheid
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathway
; TITLE OF INVENTION: create novel traits in transgenic organisms
; FILE REFERENCE: KAS-103XC1
; CURRENT APPLICATION NUMBER: US/09/918,740
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 63
; LENGTH: 8077
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon F containing A. thaliana, S. cerevisiae, and Streptomyc
; OTHER INFORMATION: CL190
; OTHER INFORMATION: DNA

US-09-918-740-63

Query Match 5.5%; Score 56.2; DB 9; Length 8077;
Best Local Similarity 45.9%; Pred. No. 2e-05;
Matches 228; Conservative 0; Mismatches 268; Indels 1; Gaps 1;

```
Qy     464 AGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGCAGATCTTCG 523
           | | | || ||  | ||  | | | | | ||    |   |  ||| | |    | |
Db    2940 AAGCGGACCTCGGCTCCTACGCCGAGCCGGTGCCCGCGGCCGACCTCGACCCGGCGCTGG 2999

Qy     524 GCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCGTCGACCCGA 583
           || | ||| | ||    |  |||     | |  | || ||||| |  || |   ||  |
Db    3000 TCATCGCCGTGGTCAACGCCGGCCCCAAGCCCGTCTCCAGCCGCGAGGCCATGCGCCGCA 3059

Qy     584 CCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGACCGTG 643
           || || | | ||  | | |      |     || | |       | | | |        |
Db    3060 CCGTCGACACCTCGCCGCTGTACCGGCCGTGGGCCGACTCCAGTAAGGACGACCTGGACG 3119

Qy     644 GTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGGCTTCA 703
           |||| | ||| | |||| || |    | |  ||||| |||     |||    ||
Db    3120 AGATGCGCTCGGCGCTGCTGCGCGGCGACCTCGAGGCCGTGGGCGAGATCGCGGAGCGCA 3179

Qy     704 ACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGTTGCCGGGCA 763
           ||| |  |  | |  | | |  || |||||| |||   ||  |     ||  | |
Db    3180 ACGCGCTCGGCATGCACGCCACCATGCTGGCCGCCCGCCCCGCGGTGCGGTACCTGTCGC 3239

Qy     764 TCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCACCCTGCGCG 823
           ||      || ||| ||||    |    |||| |  |||  || |    ||||  |
Db    3240 CGGCCACGGTCACCGTGCTCGACAGCGTGCTCCAGCTCCGCAAGGACGGTGTCCTGGCCT 3299

Qy     824 ACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTG-ATGGCGGTCACCACCGCGGGC 882
           |||  ||||     |       | |||   | | | ||   |  |     ||  || ||
Db    3300 ACGCGACCATGGACGCCGGTCCCAACGTGAAGGTGCTGTGCCGGCGGGCGGACGCCGAGC 3359

Qy     883 GGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGCCCGGTCCA 942
           |||    |  ||    || ||| |    |   ||| ||  |   |     ||| || |
Db    3360 GGGTGGCCGACGTCGTACGCGCCGCCGCGTCCGGCGGTCAGGTCCTCGTCGCCGGGCCGG 3419

Qy     943 CTGACGGTGGCCATCCG 959
           |||||||  ||  | |
Db    3420 GAGACGGTGCCCGCCTG 3436


RESULT 12
US-10-027-450-21
; Sequence 21, Application US/10027450
; Patent No. US20020102715A1
; GENERAL INFORMATION:
;  APPLICANT: Falco, Saverio Carl
;  APPLICANT:  Hitz, William D.
;  APPLICANT:  Kinney, Anthony J.
;  APPLICANT:  Cahoon, Rebecca E.
;  APPLICANT:  Rafalski, J. Antoni
;  TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC  ENZYMES
;  FILE REFERENCE: BB-1126
;  CURRENT APPLICATION NUMBER: US/10/027,450
;  CURRENT FILING DATE:  2001-12-20
;  PRIOR APPLICATION NUMBER: 60/063,423
;  PRIOR FILING DATE: 1997 October 28
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 21
;   LENGTH: 1162
;   TYPE: DNA
;   ORGANISM: Glycine max
US-10-027-450-21

  Query Match             5.4%;  Score 55.4;  DB 12;  Length 1162;
  Best Local Similarity   49.9%;  Pred. No. 3e-05;
  Matches  214;  Conservative    0;  Mismatches  191;  Indels   24;  Gaps    2;

Qy     579 CCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCAAGGA 638
           | | ||||| ||||||  | | |          |       |||   | || ||| ||||
Db     267 CGCCACCATGAAGAACGTCAACTACCTCCCCAACGTGCTGTCCATCATGGACGCCGAGGA 326

Qy     639 CCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTCCGGG 698
           |||  || |||  | |  || |   | | |||   ||   |  ||| ||||| || ||
Db     327 CCGCGGCGCGTTCGCGTCGGTGTGGGTGGACGGCGAGGGCAACGTCGCCGAGGGGCCCAT 386

Qy     699 CTTCAACGTGTGCATCGTCAAGGACGGCAA------GCTGGCCTCCCCGTCCCGGAACGC 752
           | ||||||    |||||| || || |         | ||  | |  ||| |   ||
Db     387 GGTGAACGTGGCGTTCGTCACGGCCGCCGGCGAGCTGGTGCTCCCGGCGTTCGACAAGAT 446

Qy     753 GTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGC 812
           |  | ||   ||||   |||  | || ||| |||||       | |||         |
Db     447 CCTCGCCGGGTGCACCGCCAAGCGGCTGCTCGCGCTGGCGCCGAGGCTGGTGGAGTCCGG 506

Qy     813 CACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGA------------------ 854
           |  |||    | |||||||| |||  |  ||  || || |||
Db     507 CCTCCTCAAGGCCGTCACCACCCGCCACATCGCCGCCGACGAGGCCAAGCGCTGCTCCGC 566

Qy     855 CGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGA 914
           |||| ||  |   |||  || ||     |   |||   || |||   |  |  ||| | | |
Db     567 CGAGATGGCGTTCGTCGGCAGCGGCCTCCCCGTCCTGCCCATCGTCGAGTGGGACGACCA 626

Qy     915 GGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGGGC 974
           |  | | ||| ||||  ||   ||       || | ||||  |   ||| | | ||||||
Db     627 GCTCATCGGCGACGGGAAGGTGGGGAAGACGATGATGGCGCTGTCGGATCTGCTCTGGGA 686

Qy     975 GCTGATGGA 983
           |   ||| |
Db     687 GGACATGAA 695


RESULT 13
US-09-748-033-2
; Sequence 2, Application US/09748033
; Patent No. US20020069431A1
; GENERAL INFORMATION:
;  APPLICANT: Broadway, Roxanne M.
;  APPLICANT:  Gongora, Carmenza E.
;  TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
;  TITLE OF INVENTION:  ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
;  FILE REFERENCE: 19603/3091
;  CURRENT APPLICATION NUMBER: US/09/748,033
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: 60/172,003
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1294
;   TYPE: DNA
;   ORGANISM: Streptomyces albidoflavus
US-09-748-033-2

  Query Match             5.3%;  Score 53.8;  DB 10;  Length 1294;
  Best Local Similarity   41.9%;  Pred. No. 7.7e-05;
  Matches  389;  Conservative    0;  Mismatches  537;  Indels    3;  Gaps    1;

Qy      35 TGGTGGCCGTCGAGCCCGGCGCCATCCGCGAGGACACCCCGGCCGGCTCGGTGATCCAGT 94
           ||||   |  | | |||    |   || ||       ||| ||    | |   ||||
Db     196 TGGTACGCACCTACCCCCTTCCGCACCCCGGCCGGCGCCCCTCCACGCCCGGCCTCCACC 255

Qy      95 ACAGCGACTACGAACTGGACACCTCCAGCCCGTTCGCCGGCGGCGTCGCCTGGATCGAGG 154
           ||| | |          | | ||  || | ||  || |  ||||| | || || |||
Db     256 GCAGGGGCCGGCTGACCGCCGCCCTCACCGCGGCCGTCCTCGGCGCCTCCGGGCTCGCCC 315

Qy     155 GCGAATACCTGCCGGCCGAAGAAGCGAAGATCTCCATCTTCGACACCGGATTCGGTCATT 214
           |     ||  || |||  | | |  | | || |  | | |||       |
Db     316 TCACCGGCCCCGCGACCGCCGGCGAGGGGGCCCCCGCCGCCCAGGCCGCCCCGGCCGCCG 375

Qy     215 CCGATCTGACCTACACCGTCGCGCATGTATGGCACGGCAACATCTTCCGGCTCGGCGACC 274
                   |          ||   |      | ||  ||||| |  | |  ||   |
Db     376 TACCGGCCCACGCGGTGACCGGTTACTGGCAGAACTTCAACAACGGCGCGACCGTGCAGA 435

Qy     275 ACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGCCGGGTACAGCAAGG 334
           ||| | |    |||  ||||       ||   |  | |   |  ||   ||  || ||
Db     436 CCCTCGCCGACGTGCCGGACGCCTACGACATCATCGCCGTCTCCTTCGCCGACGCCACGG 495

Qy     335 ACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCTGCGCGAATCGTTCG 394
           | |   || ||||||||||   |    ||  ||| |   ||      || |  |   ||
```

```
Db    496 CCAACGCGGGCGAGATCACCTTCACCCTCGACTCGGTCGGGCTCGGCGGCTACACCGACG 555

Qy    395 TGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAAGGACCTGTCCAAGC 454

Db    556 AGCAGTTCCGCGCCGACCTCGCCGCCAAGCAGGCCGACGGCAAGTCGGTGATCATCTCGG 615

Qy    455 TCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTTCCCGCCCGCCGAGC 514

Db    616 TCGGCGGCGAGAAGGGCGCGGTCGCCGTCAACGACAGCGCCTCCGCCCAGCGCTTCGCCG 675

Qy    515 AGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGCCGGCCGCAACACCG 574

Db    676 ACAGCACCTACGCGCT---GATGGAGGAGTACGGCTTCGACGGCGTCGACATCGACCTGG 732

Qy    575 TCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGCCAGTTTCGAAGCCA 634

Db    733 AGAACGGCCTCAACTCCACCTACATGACCGAGGCCCTCACCAAGCTCCACGAGAAGGCCG 792

Qy    635 AGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTGCGTGGCCGAAGGTC 694

Db    793 GGGACGGCCTGGTCCTCACCATGGCGCCGCAGACCATCGACATGCAGTCGCCCGAGAACG 852

Qy    695 CGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCCGTCCCGGAACGCGT 754

Db    853 AGTACTTCAAGACGGCGCTGGTCACGAAAGACTTCCTGACCGCCGTCAACATGCAGTACT 912

Qy    755 TGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGATGGGCATCGAAGCCA 814

Db    913 ACAACAGCGGCTCGATGCTCGGCTGCGACGGCCAGGTCTACGCGCAGGGCACCGTCGACT 972

Qy    815 CCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTTGATGGCGGTCACCA 874

Db    973 TCCTCACCGCGCTCGCCTGCATCCAGCTGGAGAACGGTCTCGACGCCTCCCAGGTCGGCA 1032

Qy    875 CCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGTGGGCAACGGCGAGC 934

Db   1033 TCGGTGTCCCCGCCTCCCCGAAGGCGGCCGGCGGCGGCTACGTCGAGCCCTCCGTGGTCA 1092

Qy    935 CCGGTCCACTGACGGTGGCCATCCGGGAC 963

Db   1093 ACGACGCGCTGGACTGCCTGACCCGGGGC 1121


RESULT 14
US-09-974-300-492
; Sequence 492, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 492
;   LENGTH: 810
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-492

  Query Match             5.1%;  Score 52.4;  DB 10;  Length 810;
  Best Local Similarity   48.4%;  Pred. No. 0.00017;
  Matches  180;  Conservative    0;  Mismatches  186;  Indels    6;  Gaps    1;

Qy    619 GCCAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAAC 678

Db    420 GTCCGGATCGAGGCGCATATGGCGGGTGTGACGGAAGCGCTCATGCTCAATGATCAAGGC 479

Qy    679 TGCGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCC 738

Db    480 TATGTCGCCGAAGGGTCTGCGGATAACGTATTTATTTATAAAAACGGAAAGCTCTTGACG 539

Qy    739 CCGTCCCGGAA------CGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCC 792

Db    540 CCTCCGGGCTATATCGGAGCGCTTGAAGGAATCACCCGGAATGCCATCATCGAAATAGCG 599

Qy    793 GACCAGATGGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCC 852

Db    600 CGAGAGCTCGGCTATGAAGTGAAAGAAGAGCCGTTTACCCGCCATGACGTATACACAGCC 659

Qy    853 GACGAGTTGATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGC 912

Db    660 GAGGAAGTGTTTTTAACCGGAACGGCTGCAGAAGTCATCGCGGTCGTAAAAGTTGACGGC 719

Qy    913 GAGGCCGTGGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTCTGG 972

Db    720 CGCAAGATCGGGGACGGCAAACCGGGAGTCCACACAAACCGGATGCTTGAAAAGTTCCGC 779

Qy    973 GCGCTGATGGAC 984

Db    780 GAGCGCGTCGTC 791


RESULT 15
US-10-145-415-98
; Sequence 98, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
;  APPLICANT: Buckel, Thomas Gunter
;  APPLICANT:  Hammer, Philip Eugene
;  APPLICANT:  Hill, Dwight Steven
;  APPLICANT:  Ligon, James Madison
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Pachlatko, Johannes Paul
;  APPLICANT:  Zirkle, Ross Eric
;  TITLE OF INVENTION: Methods and Compositions for Making Emamectin
;  FILE REFERENCE: SYN-117 109846.312
;  CURRENT APPLICATION NUMBER: US/10/145,415
;  CURRENT FILING DATE:  2002-05-14
;  PRIOR APPLICATION NUMBER: US 60/291,149
;  PRIOR FILING DATE: 2001-05-16
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 98
;   LENGTH: 1266
;   TYPE: DNA
;   ORGANISM: Streptomyces
US-10-145-415-98

  Query Match             5.0%;  Score 51;  DB 9;  Length 1266;
  Best Local Similarity   43.9%;  Pred. No. 0.0004;
  Matches  311;  Conservative    0;  Mismatches  395;  Indels    3;  Gaps    2;

Qy    261 CCGGCTCGGCGACCACCTGGACCGGTTGCTCGACGGGGCGTCCAAGCTGCGCCTGGACGC 320

Db    207 CTGGTACGCCCAGGCACAGATCGAACTGCACCTGGGCCAGCCCGCCGTCCGCCTCGACCC 266

Qy    321 CGGGTACAGCAAGGACGAACTGGCCGAGATCACCAAGAAGTGCGTGTCGATGTCGCAGCT 380

Db    267 CGAGGGCAGGACCGTCCGCCTCGGCGACGGCACCCTGATCGCCTACGACAAGCTGCTGCT 326

Qy    381 GCGCGAATCGTTCGTGAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGGGCGAGAA 440

Db    327 GGCCACCGGCGCCGAACCGCGGCGCCTGGACATCCCCGGCACCGGCCTGGCCGGCGTGCA 386

Qy    441 GGACCTGTCCAAGCTCACCCATCAGGTGTACATCTACGCCATCCCGTACCTGTGGGCCTT 500

Db    387 CCACCTGCGCCGCCTCGCCCACGCCGAACGGCTGCGCGGCGTCCTGGCCTCCCTCGGCCG 446
```

```
Qy     501 CCCGCCCGCCGAGCAGATCTTCGGCACCACCGCGATCGTGCCGCGCCATGTCCGCCGCGC 560
           |     || | | | | |'|  ||| |    |||      |    |   | |||  |  |||
Db     447 CGACAACGGCCATCTGGTGATCGCCGGAGCCGGCTGGATCGGCCTGGAGGTCGCCGCCGC 506

Qy     561 CGGCCGCAACACCGTCGACCCGACCATCAAGAACTACCAGTGGGGTGATCTCACCGCAGC 620
            | ||||  |  || || |  |   | |     | |              |       |
Db     507 GGCCCGCTCCTACGGCGCCGAGGTGACCGTCGTCGAGGCCGCCCCGACGCCGCTGCACGG 566

Qy     621 CAGTTTCGAAGCCAAGGACCGTGGTGCGCGCACCGCGATCCTGCTCGACTCGGACAACTG 680
           ||   | |   || |   | | |||  |  |||||   | |  | |||    | |  | |
Db     567 CATCCTGGGGCCCGAACTCGGCGGTCTGTTCACCGATCTGCACCGCGAGCACGGCGTCCG 626

Qy     681 CGTGGCCGAAGGTCCGGGCTTCAACGTGTGCATCGTCAAGGACGGCAAGCTGGCCTCCCC 740
           | |   |   ||  |  |||||| || |  | |||   ||| ||||| | ||  ||| ||
Db     627 CTTCCACTTCGGCGCCCGCTTCACCGAGATCGTCGGAGAGGGCGGCATGGTG--CTCGCC 684

Qy     741 GTCCCGGAACGCGTTGCCGGGCATCACCCGTAAGACGGTGTTCGAACTGGCCGACCAGAT 800
           || |       ||  | || | | |||||     ||| ||| |   | |         |
Db     685 GTGCGCACCGACGACGGCGAGGAACACCCCGCCCACGATGTGCTCGCCGCGATCGGCGCC 744

Qy     801 GGGCATCGAAGCCACCCTGCGCGACGTCACCAGCCGTGAACTCTACGACGCCGACGAGTT 860
           |  |  || | || | ||   |||     || | |  || |||  |||| | ||
Db     745 GCCCCGCGCA-CCGCGCTCGCCGAACAGGCCGGGCTGGATCTCGCCGACCCGGAGACCGG 803

Qy     861 GATGGCGGTCACCACCGCGGGCGGGGTCACACCGATCAACTCGCTGGATGGCGAGGCCGT 920
           | |||  ||  ||  |  | | |     |   | || ||       ||  ||||
Db     804 CGGCGGGGTGGCCGTCGACGCGGCGCTGCGCACCTCCGACCCGTACATCTACGCCGCCGG 863

Qy     921 GGGCAACGGCGAGCCCGGTCCACTGACGGTGGCCATCCGGGACCGGTTC 969
           | |  || ||   |||  |  | |  | ||| || ||||   ||| ||
Db     864 TGACGTCGCCGCCGCCGACCACCCGCTGCTGGACACCCGGCTGCGGGTC 912
```

Search completed: April 24, 2003, 22:03:56
Job time : 156 secs